

STIC-Biotech/ChemLib

79512

From: Davis, Minh-Tam
Sent: Tuesday, November 05, 2002 4:18 PM
To: STIC-Biotech/ChemLib
Subject: Search request for 09/899569

SEQ ID NO: 3 and 4
Thank you.
MINH TAM DAVIS
ART UNIT 1642, ROOM 8A01, MB 8E12
305-2008

Mary Jane Ruhl
Tech. Info. Specialist, STIC
TC-1600
CM-1, Room 6A-06
Phone: 605-1155

RECEIVED
NOV - 5 2002
STIC

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: _____
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

This Page Blank (uspto)

PD 17-JAN-2002.
 XX
 PF 05-JUL-2001; 2001MO-EP07705.
 XX
 PR 07-JUL-2000; 2000DE-1033080.
 PR 19-APR-2001; 2001DE-1019294.
 XX
 PA (BOEH) BOEHRINGER INGELHEIM INT GMBH.
 PI Schwieler N, Scherl-Mostageer M, Sommergruber W, Abseher R:
 XX WPI: 2002-171704/22.
 DR P-PSDB; AAM49641.
 XX
 PT New tumor-associated antigen B345, useful for diagnosis and
 immunotherapy of tumors, also related nucleic acid and antibodies
 PS Claim 3; Page 79-85; 102pp; German.
 XX
 CC This invention describes a novel tumour-associated antigen, designated
 CC B345 which has cytostatic activity. B345 is involved in communication,
 CC interaction and/or signal transduction with extracellular components and
 CC ligands, especially in the metastatic potential of cancers, particularly
 CC of the colon. B345 or its immunogenic fragments, also the DNA that
 CC encodes it, are useful for immunotherapy of cancer, particularly
 CC carcinoma of lung or colon. Antibodies raised against B345 are useful for
 CC treatment and diagnosis of cancers that are associated with B345
 CC expression, including their use for targeted delivery of cytotoxic or
 CC radioactive agents. Probes derived from B345 can be used to detect
 CC tumour-specific mutations in the B345 sequence, and can be used to screen
 CC for B345 specific modulators. This sequence encodes a human B345
 CC tumour-associated antigen described in the invention.
 CC
 XX
 SQ Sequence 6163 BP; 1501 A; 1689 C; 1482 G; 1491 T; 0 other;
 Query Match 100.0%; Score 6163; DB 24; Length 6163;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 6163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 481 ATGTGTTCATCAAGTCTGAGAAAGAAATAGTCTTTACCTTTAGCTCCAGAGTCTCGAG 540
 Qy 541 AATCACTTTGTCATAGAGATCCAGAAAAATATTGACTGTATGTCAGGCCCATGCTCTTT 600
 Db 541 AATCACTTTGTCATAGAGATCCAGAAAAATATTGACTGTATGTCAGGCCCATGCTCTTT 600
 Qy 601 GGGAGGTTCAAGCTTCAAGCCCTGACATGCTGTGCTTACCTCAACAGAACTTTATC 660
 Db 601 GGGAGGTTCAAGCTTCAAGCCCTGACATGCTGTGCTTACCTCAACAGAACTTTATC 660
 Qy 661 TGGGATGTCAAAGCTCAAGAGATGAGTGTGAGCTGAGCTTTTCATCCCTCGGCTG 720
 Db 661 TGGGATGTCAAAGCTCAAGAGATGAGTGTGAGCTGAGCTTTTCATCCCTCGGCTG 720
 Qy 721 AGGAGATCCGCTCGGCTGAGAGCTGCGCAGAGAGATCACTCACTCATCAGGCGGA 780
 Db 721 AGGAGATCCGCTCGGCTGAGAGCTGCGCAGAGAGATCACTCACTCATCAGGCGGA 780
 Qy 781 ATGATGTCACCGTGTGTCAGATCGAATCTTCTGACGAATGGCACTGTCCGGATC 840
 Db 781 ATGATGTCACCGTGTGTCAGATCGAATCTTCTGACGAATGGCACTGTCCGGATC 840
 Qy 841 AAGATGCAAGAGAGTGAATAAGGCTTACCTCCATGCTTCCACCCAGAAATGTC 900
 Db 841 AAGATGCAAGAGAGTGAATAAGGCTTACCTCCATGCTTCCACCCAGAAATGTC 900
 Qy 901 TCCGCTTTCAGCATTTGCAAAACCGCTCATCTATAAAAGCTGTGTCATCATGATCTGTG 960
 Db 901 TCCGCTTTCAGCATTTGCAAAACCGCTCATCTATAAAAGCTGTGTCATCATGATCTGTG 960
 Qy 961 TTTGAGGTTGAAGGCTGAGCAACCTGATGTCTGCAACTTACCAAGAGCTTCCCTGAG 1020
 Db 961 TTTGAGGTTGAAGGCTGAGCAACCTGATGTCTGCAACTTACCAAGAGCTTCCCTGAG 1020
 Qy 1021 GATGAGCTCATGAGCTGGAGTTTGTGCTCTGACACACTGGGCGGAGGCTGCTTTC 1080
 Db 1021 GATGAGCTCATGAGCTGGAGTTTGTGCTCTGACACACTGGGCGGAGGCTGCTTTC 1080
 Qy 1081 CTCAACTTCAACCTCTTCAACTGTGAGAGAGAGAGGCGGGTTGAATACTACATCCG 1140
 Db 1081 CTCAACTTCAACCTCTTCAACTGTGAGAGAGAGAGGCGGGTTGAATACTACATCCG 1140
 Qy 1141 GGTCTCCACCAACCAACCCGAGGTGTTCAAGCTGAGAGACAAGAGCTGGGAACATGGCG 1200
 Db 1141 GGTCTCCACCAACCAACCCGAGGTGTTCAAGCTGAGAGACAAGAGCTGGGAACATGGCG 1200
 Qy 1201 GGGAACTTCAACCTCTTCTGCAAGGCTGTGACCAAGATGCCAAATCTCAGGATCTCT 1260
 Db 1201 GGGAACTTCAACCTCTTCTGCAAGGCTGTGACCAAGATGCCAAATCTCAGGATCTCT 1260
 Qy 1261 CGGCTGCAGTTCCAAAGTTTGTGTCAAATCTCAACAAATGAAGCAATTAATCTACGTG 1320
 Db 1261 CGGCTGCAGTTCCAAAGTTTGTGTCAAATCTCAACAAATGAAGCAATTAATCTACGTG 1320
 Qy 1321 GTTGACTTGAATATGAGCAGGATGTCACTCACTCAGGACACGCGCTCAACAG 1380
 Db 1321 GTTGACTTGAATATGAGCAGGATGTCACTCACTCAGGACACGCGCTCAACAG 1380
 Qy 1381 AGCGGCAAGTTTGCCCTGCTGTTGCTGATGATCAATCTGAGACCTGAGTAGCAAC 1440
 Db 1381 AGCGGCAAGTTTGCCCTGCTGTTGCTGATGATCAATCTGAGACCTGAGTAGCAAC 1440
 Qy 1441 CTCACCTGACATCTGGCTCCAAACACAAATCTCTCTTGTGATGATCTGACAGT 1500
 Db 1441 CTCACCTGACATCTGGCTCCAAACACAAATCTCTCTTGTGATGATCTGACAGT 1500
 Qy 1501 CTGTGATGATGATGAGAAAAACCAATAGCTGACAGACACGCGTACTGCAAAAGAAA 1560
 Db 1501 CTGTGATGATGATGAGAAAAACCAATAGCTGACAGACACGCGTACTGCAAAAGAAA 1560
 Qy 1561 TCTTACTCATCTCAGAGTCCCAATGATCACTCTGCTGTGAGAGTGCATGATCTTC 1620
 Db 1561 TCTTACTCATCTCAGAGTCCCAATGATCACTCTGCTGTGAGAGTGCATGATCTTC 1620

OY	1621	TCCTGGAAGCTGCTGGTGTGCCCAAGAGACAGGCTACGCTTGCTGTGTGCAAGCCAGAG	1680
Dp	1621	TCCCTGGGAAGCTGCTGGTGTGCCCAAGAGACAGGCTACGCTTGCTGTGTGCAAGCCAGAG	1680
OY	1681	CTGGAAGACGATACACACGAGAAAGCCCTGCAACACAGCTTAGAGTACCTGCTGGCCAGT	1740
Dp	1681	CTGGAAGACGATACACACGAGAAAGCCCTGCAACACAGCTTAGAGTACCTGCTGGCCAGT	1740
OY	1741	GCCATTACCCAGCCAGGACGCTGTAACCTTCGGCTCCTTTGCCCCGGAGGGCTATCAAGAG	1800
Dp	1741	GCCATTACCCAGCCAGGACGCTGTAACCTTCGGCTCCTTTGCCCCGGAGGGCTATCAAGAG	1800
OY	1801	ATCCAGGTGAAGCAGAAACATCTCGGTGACCTTCGACACTTTGCCCCAGCTTCCACA	1860
Dp	1801	ATCCAGGTGAAGCAGAAACATCTCGGTGACCTTCGACACTTTGCCCCAGCTTCCACA	1860
OY	1861	GAGGCTTCACAGGCGAGGTGTGACGGTGTCTTTATACCTATTATTTCAAAGAGAGGCTT	1920
Dp	1861	GAGGCTTCACAGGCGAGGTGTGACGGTGTCTTTATACCTATTATTTCAAAGAGAGGCTT	1920
OY	1921	TTACAGGTAAGCCCTTGACACAAAAAGCAAGGTGTACTCTAGAGACCCCAACTGGGAGCGG	1980
Dp	1921	TTACAGGTAAGCCCTTGACACAAAAAGCAAGGTGTACTCTAGAGACCCCAACTGGGAGCGG	1980
OY	1981	GGCCTGCATCCCTCACCTCTGTGTCTGTGAAACATCAGCGTGCACGAGACAGGTGGCC	2040
Dp	1981	GGCCTGCATCCCTCACCTCTGTGTCTGTGAAACATCAGCGTGCACGAGACAGGTGGCC	2040
OY	2041	TGCTGACTTCTTTAAAGAGCGGAGCGGCGTGTGTGCGACAGACAGGCGCGCATTCATG	2100
Dp	2041	TGCTGACTTCTTTAAAGAGCGGAGCGGCGTGTGTGCGACAGACAGGCGCGCATTCATG	2100
OY	2101	ATCATCCAGGAGCGAGCGGACCGGGCGTGAAGATTTTAGCGTGAAGAGATGTGTC	2160
Dp	2101	ATCATCCAGGAGCGAGCGGACCGGGCGTGAAGATTTTAGCGTGAAGAGATGTGTC	2160
OY	2161	CCCAAGCCAAAGCTTCCACATCAACAGCTTCTGGGTCAACATCTCTAACGACGCCACAG	2220
Dp	2161	CCCAAGCCAAAGCTTCCACATCAACAGCTTCTGGGTCAACATCTCTAACGACGCCACAG	2220
OY	2221	AGCGGCAAGACGCTAGACCTGTCTTCTGGGTGACACTTAACCCCAAGGAGCTGTGACTTG	2280
Dp	2221	AGCGGCAAGACGCTAGACCTGTCTTCTGGGTGACACTTAACCCCAAGGAGCTGTGACTTG	2280
OY	2281	ACTGTCATCCATCGGAGGCGGTGGGAGGAGTGTACTGCTGTGTGCTGTGCGCTCGGCTC	2340
Dp	2281	ACTGTCATCCATCGGAGGCGGTGGGAGGAGTGTACTGCTGTGTGCTGTGCGCTCGGCTC	2340
OY	2341	ATCATTTGCTGTGTGAAGAAAGAAAGAAAGAACAAACAAAGGCGCGCTGTGGGTATC	2400
Dp	2341	ATCATTTGCTGTGTGAAGAAAGAAAGAAAGAAAGAACAAACAAAGGCGCGCTGTGGGTATC	2400
OY	2401	TACAAATGGCAACATCAATACTGAGATGCGGAGCGACCCAAAAAAGTTTCAGAAAGGCGGA	2460
Dp	2401	TACAAATGGCAACATCAATACTGAGATGCGGAGCGACCCAAAAAAGTTTCAGAAAGGCGGA	2460
OY	2461	AAGGACAATGATCCCATGTGTATGCAAGTCATGAGAGAACACATGATATATGGGATCTG	2520
Dp	2461	AAGGACAATGATCCCATGTGTATGCAAGTCATGAGAGAACACATGATATATGGGATCTG	2520
OY	2521	CTACAGGATTCACAGCGGCTCTTCTCTGACAGCCAGAGGTGGAACCTACCGGCGCTTCAG	2580
Dp	2521	CTACAGGATTCACAGCGGCTCTTCTCTGACAGCCAGAGGTGGAACCTACCGGCGCTTCAG	2580
OY	2581	GGACACATGGGGGTCTGTCTCCTCCGCCACCCACCATATGCTCCAGGCGCCCAACTGCA	2640
Dp	2581	GGACACATGGGGGTCTGTCTCCTCCGCCACCCACCATATGCTCCAGGCGCCCAACTGCA	2640
OY	2641	AAGTTGGCACTAGAGAGCAACTCTCGCTCCCTCTCTGACTGTGAGAGTGAACCGTAC	2700
Dp	2641	AAGTTGGCACTAGAGAGCAACTCTCGCTCCCTCTCTGACTGTGAGAGTGAACCGTAC	2700

QY	2701	ACCTCTCCCAATCCCAACAATTTGGGAGATGTAAAGCAGCAAGAGACAGACATTTCCCTTACTG	2760
Db	2701	ACCTCTCTCCCAATCCCAACAATTTGGGAGATGTAAAGCAGCAAGAGACAGACATTTCCCTTACTG	2760
QY	2761	AACACTCAGAGCCCATGAGGCCAGCAGAAATTAATCTTGATATCCATTCAGACGCTTGTGCTGA	2820
Db	2761	AACACTCAGAGCCCATGAGGCCAGCAGAAATTAATCTTGATATCCATTCAGACGCTTGTGCTGA	2820
QY	2821	GTTTTCATTAAGCAGGAGACAGAGACACCCGCTCCGTTCTCTTAACAGAGAAATCTTAAGAA	2880
Db	2821	GTTTTCATTAAGCAGGAGACAGAGACACCCGCTCCGTTCTCTTAACAGAGAAATCTTAAGAA	2880
QY	2881	GAGGAATTAATACAGAGAGAACAGCAGAGAGTTTTCCTGACACCGCCAACTTCACATTCG	2940
Db	2881	GAGGAATTAATACAGAGAGAACAGCAGAGAGTTTTCCTGACACCGCCAACTTCACATTCG	2940
QY	2941	TCAGAGACCTCATCTTAAGGAGCAAGACATGAAATGAATGAATTCATTCGATGATACAGT	3000
Db	2941	TCAGAGACCTCATCTTAAGGAGCAAGACATGAAATGAATGAATTCATTCGATGATACAGT	3000
QY	3001	CATGACACCTCATCTGCTCCCAACTTAGGCTGTGGGTAGCCAGCCGCTAATGAGAG	3060
Db	3001	CATGACACCTCATCTGCTCCCAACTTAGGCTGTGGGTAGCCAGCCGCTAATGAGAG	3060
QY	3061	AGAGAGCCCTGATGTCACCTAGACATAGGGTTGACAGAACCCCTGATTCAGAGTGTAAAC	3120
Db	3061	AGAGAGCCCTGATGTCACCTAGACATAGGGTTGACAGAACCCCTGATTCAGAGTGTAAAC	3120
QY	3121	AGAGGCTTGCCCTCTTACAGAGCAACAGTTCCATTTCCAGAGAGCCTACTGAGGTCCCTA	3180
Db	3121	AGAGGCTTGCCCTCTTACAGAGCAACAGTTCCATTTCCAGAGAGCCTACTGAGGTCCCTA	3180
QY	3181	CTCTCACTGGGGGTCCCAAGATGAAACGACAAATGTGCTTTTATATATATATTTTGG	3240
Db	3181	CTCTCACTGGGGGTCCCAAGATGAAACGACAAATGTGCTTTTATATATATATTTTGG	3240
QY	3241	TGCTCTGTGTATTTAAGAGATCAAAATGTATTAACACCTAGCTCTTTACCTGACTTA	3300
Db	3241	TGCTCTGTGTATTTAAGAGATCAAAATGTATTAACACCTAGCTCTTTACCTGACTTA	3300
QY	3301	GTAATACTCTACTAATCTGTTTGGATGCTCGGGTTTGAGCTTCTACTGACGCCGTAAAT	3360
Db	3301	GTAATACTCTACTAATCTGTTTGGATGCTCGGGTTTGAGCTTCTACTGACGCCGTAAAT	3360
QY	3361	AAACGTGCGCTGTGCCCCCAGAGGTGGTGAATTAATTAATCTGCAACAGAGAAAAGA	3420
Db	3361	AAACGTGCGCTGTGCCCCCAGAGGTGGTGAATTAATTAATCTGCAACAGAGAAAAGA	3420
QY	3421	ATGTGTGTGTTTGAGCAGACATGTGACACATATGCTTTGATAAGAGACTTCTGATTCYC	3480
Db	3421	ATGTGTGTGTTTGAGCAGACATGTGACACATATGCTTTGATAAGAGACTTCTGATTCYC	3480
QY	3481	TAGGTCGGTTCGTGGTTATCCCATTTGGGAAATTCATCTTGAATCCCATTTGCTCATATGT	3540
Db	3481	TAGGTCGGTTCGTGGTTATCCCATTTGGGAAATTCATCTTGAATCCCATTTGCTCATATGT	3540
QY	3541	CCTAGCAATTAAGAGAAATTCCTCAATGTCATGTCGAGGTTCCTCTAGCTCAGACATAA	3600
Db	3541	CCTAGCAATTAAGAGAAATTCCTCAATGTCATGTCGAGGTTCCTCTAGCTCAGACATAA	3600
QY	3601	CTTTGACATTTAAGAGAAATTTAGAGAAATTTCTCATCTCTAATAAAATGTTTAATATA	3660
Db	3601	CTTTGACATTTAAGAGAAATTTAGAGAAATTTCTCATCTCTAATAAAATGTTTAATATA	3660
QY	3661	TACCAAAAGTGGCCCCCTGCTAGTTTCTGTTCGTGACACCTGATTAATGCTG	3720
Db	3661	TACCAAAAGTGGCCCCCTGCTAGTTTCTGTTCGTGACACCTGATTAATGCTG	3720
QY	3721	CTTAATAAACACACATTAAGCTTAAGTCGTGGGGAATCAGAAATTCGAAATGGATGCTCT	3780
Db	3721	CTTAATAAACACACATTAAGCTTAAGTCGTGGGGAATCAGAAATTCGAAATGGATGCTCT	3780
QY	3781	GATGTAATAATCAAGTGTGACAGAGCTGTGCTCTTTCGAAGGCTCTAGGGAGAGACCG	3840

Db 3781 GAAAGAAATCAAGGTGTCAGACAGAGCTGTGCTCTTCTGAAAGCTCTAGGAGAAACCG 3840
QY 3841 GTTCTTGCCATTTCAAGCTCTAGAGCTGGCTGCATTTCCAGGCTCCAGTGGCTGTC 3900
Db 3841 GTTCTTGCCATTTCAAGCTCTAGAGCTGGCTGCATTTCCAGGCTCCAGTGGCTGTC 3900
QY 3901 AAGCTTTCTCACATGGCATCTGTGACACTGGCCCTCCACATCTCTTTGACATTC 3960
Db 3901 AAGCTTTCTCACATGGCATCTGTGACACTGGCCCTCCACATCTCTTTGACATTC 3960
QY 3961 AAAGCCCAACGAGAAATCCAGATTCCTCATCTAAAGATCTTCAATCCCTGGA 4020
Db 3961 AAAGCCCAACGAGAAATCCAGATTCCTCATCTAAAGATCTTCAATCCCTGGA 4020
QY 4021 AGAGCTTTTGCCATGCAAGACATAGCCACAGTGGGGATTAGGACAGGACATCTT 4080
Db 4021 AGAGCTTTTGCCATGCAAGACATAGCCACAGTGGGGATTAGGACAGGACATCTT 4080
QY 4081 TGGGGTCTGTATTCTGCTTACCAACACTTCTGCACTGATCCACAGAGAGGCTA 4140
Db 4081 TGGGGTCTGTATTCTGCTTACCAACACTTCTGCACTGATCCACAGAGAGGCTA 4140
QY 4141 CAAATGATCTGGGCAAGGAGATTTTGTAGCTTGGGAGCTTAACCTTAAAAA 4200
Db 4141 CAAATGATCTGGGCAAGGAGATTTTGTAGCTTGGGAGCTTAACCTTAAAAA 4200
QY 4201 ACCCCAGATCAGAAATCTGGCCATGCTGGGGCTCACATTTCCACTAGCAACACTGGC 4260
Db 4201 ACCCCAGATCAGAAATCTGGCCATGCTGGGGCTCACATTTCCACTAGCAACACTGGC 4260
QY 4261 TGGAGCTGGGCAACAGCTCTGCTTTAGAGGGGTGTCCACTTCCAGAGTCCACAGC 4320
Db 4261 TGGAGCTGGGCAACAGCTCTGCTTTAGAGGGGTGTCCACTTCCAGAGTCCACAGC 4320
QY 4321 CCACACTACGCTTATCACTTCCACATAGAGGCTAGTGTCTTCTACTGATCATG 4380
Db 4321 CCACACTACGCTTATCACTTCCACATAGAGGCTAGTGTCTTCTACTGATCATG 4380
QY 4381 CCCCTGCAAGTTGCAATTTATGATGAAAAAGAAAGACTGGATTAATCTCTAATCAG 4440
Db 4381 CCCCTGCAAGTTGCAATTTATGATGAAAAAGAAAGACTGGATTAATCTCTAATCAG 4440
QY 4441 TGAATACCATGAGACCAATGTGTCTCACATTTACCTTTTCTTTTCTTTTCT 4500
Db 4441 TGAATACCATGAGACCAATGTGTCTCACATTTACCTTTTCTTTTCTTTTCT 4500
QY 4501 TTTTCTTTTCTTTTATGAGAGAGATCTCATCTGTGTGCTAGGCTGGAGTGCAG 4560
Db 4501 TTTTCTTTTCTTTTATGAGAGAGATCTCATCTGTGTGTGCTAGGCTGGAGTGCAG 4560
QY 4561 TGGGCAATCTCGGCTCACTGCAACCTGTGCTGGGCTCAAGCAATTTCCACACTC 4620
Db 4561 TGGGCAATCTCGGCTCACTGCAACCTGTGCTGGGCTCAAGCAATTTCCACACTC 4620
QY 4621 AGCTTCCCAATAGCTGGGATCACTGGCAACCAACCATGCCCCAGCTAATTTGTATTT 4680
Db 4621 AGCTTCCCAATAGCTGGGATCACTGGCAACCAACCATGCCCCAGCTAATTTGTATTT 4680
QY 4681 TTTGTAGAGAGGCTTTCAACATGTTGCCAGGCTGTCTCAACCTCTGGGCTCAAGC 4740
Db 4681 TTTGTAGAGAGGCTTTCAACATGTTGCCAGGCTGTCTCAACCTCTGGGCTCAAGC 4740
QY 4741 AATCTCTGCTGCTGGCTCCCAAAAGTGTGGATTAAGATGAGCCACCGCATCTCAG 4800
Db 4741 AATCTCTGCTGCTGGCTCCCAAAAGTGTGGATTAAGATGAGCCACCGCATCTCAG 4800
QY 4801 CCCACACCTCATTTATACCAATTTACCTGCCAGTAAGTGTGACTTTGCTCTCAC 4860
Db 4801 CCCACACCTCATTTATACCAATTTACCTGCCAGTAAGTGTGACTTTGCTCTCAC 4860
QY 4861 CCTGCTCTGATCTGGAAGAGAGGATTTATGCTTGTAGACAGTCCCAAGTT 4920
Db 4861 CCTGCTCTGATCTGGAAGAGAGGATTTATGCTTGTAGACAGTCCCAAGTT 4920

Db 4861 CCTGCTCTGATCTGGAAGAGAGGATTTATGCTTGTAGACAGTCCCAAGTT 4920
QY 4921 CAATATTTCTGGCGCAAAAATCTTCAAAAATAATGACTTCAATTTGATTTCAATGA 4980
Db 4921 CAATATTTCTGGCGCAAAAATCTTCAAAAATAATGACTTCAATTTGATTTCAATGA 4980
QY 4981 ATTCACTTGGAAATGCAACCGCTCAACCTGTTTACATGGCATTAATGAAGAATTTTA 5040
Db 4981 ATTCACTTGGAAATGCAACCGCTCAACCTGTTTACATGGCATTAATGAAGAATTTTA 5040
QY 5041 TAGTCTCTAAATGGGCTGATCTGCAAGACCTCTGGAACCTTTCCAGAGATGATAT 5100
Db 5041 TAGTCTCTAAATGGGCTGATCTGCAAGACCTCTGGAACCTTTCCAGAGATGATAT 5100
QY 5101 TTAAGTATGCTTGGCGTTGCTTATGGAACCTTTCCCTTCTGAAAGTCTGTTCTGTC 5160
Db 5101 TTAAGTATGCTTGGCGTTGCTTATGGAACCTTTCCCTTCTGAAAGTCTGTTCTGTC 5160
QY 5161 CCACTGACCTTGGCTTGTGAGCGAGATGCTGACCTTGATTAAGGGCCAAAGAGG 5220
Db 5161 CCACTGACCTTGGCTTGTGAGCGAGATGCTGACCTTGATTAAGGGCCAAAGAGG 5220
QY 5221 CTGGGCTCTCTTCCCTCAGTGAAGAGCCCTATTGAATTCATGTTGGAGCCCTAGC 5280
Db 5221 CTGGGCTCTCTTCCCTCAGTGAAGAGCCCTATTGAATTCATGTTGGAGCCCTAGC 5280
QY 5281 CCTTCATTTCTGACATTTCCCAACCTCCAGACCCCTTCCAAAGAGGACTAGGCTCTGC 5340
Db 5281 CCTTCATTTCTGACATTTCCCAACCTCCAGACCCCTTCCAAAGAGGACTAGGCTCTGC 5340
QY 5341 ATTCCACCAAAGTGGGATTTGGCTTCTTAAGCTGGCTTCTTACCATCACCGACAT 5400
Db 5341 ATTCCACCAAAGTGGGATTTGGCTTCTTAAGCTGGCTTCTTACCATCACCGACAT 5400
QY 5401 CACTGTTGCCCTGCAAGGACACACAGTGGCCATTTCTTCAACAGAGGGCTCAAACTCC 5460
Db 5401 CACTGTTGCCCTGCAAGGACACACAGTGGCCATTTCTTCAACAGAGGGCTCAAACTCC 5460
QY 5461 TGGCAAGTGTCTGGCTCTGAGACAGTATTTCTGAGGCTGTGCTCAGTGAAGGGC 5520
Db 5461 TGGCAAGTGTGTGGCTCTGAGACAGTATTTCTGAGGCTGTGCTCAGTGAAGGGC 5520
QY 5521 CCAGCTGAGGAACCTTGGCTCTTTTCTTAAAGCCAGGCCCCACTTACATTAACATTT 5580
Db 5521 CCAGCTGAGGAACCTTGGCTCTTTTCTTAAAGCCAGGCCCCACTTACATTAACATTT 5580
QY 5581 TCAAGGTCACGTGGAACAGTGAAGTGCATTTGTTGAAGCTATGATGACAGCCACT 5640
Db 5581 TCAAGGTCACGTGGAACAGTGAAGTGCATTTGTTGAAGCTATGATGACAGCCACT 5640
QY 5641 GCTCATCCAGTGGTGTGCCATGCTTACGAGAAAGGCCAGGSCATGAGGACTGGCTCT 5700
Db 5641 GCTCATCCAGTGGTGTGCCATGCTTACGAGAAAGGCCAGGSCATGAGGACTGGCTCT 5700
QY 5701 AATGCTGTGCTATTTGACAGAGAGGAAAGGTCTCAAGAAAGTCAACTGAGCAAGCA 5760
Db 5701 AATGCTGTGCTATTTGACAGAGAGGAAAGGTCTCAAGAAAGTCAACTGAGCAAGCA 5760
QY 5761 CAAGCCACCGGACATGGGCTTGGTAAAGTTAGCAACAGTGTGTGGATTTGAGT 5820
Db 5761 CAAGCCACCGGACATGGGCTTGGTAAAGTTAGCAACAGTGTGTGGATTTGAGT 5820
QY 5821 GCTTCACTGGAATTAATTTATTTATTTGACATATTTTATAGTGGCATTTTATCATTTTC 5880
Db 5821 GCTTCACTGGAATTAATTTATTTATTTGACATATTTTATAGTGGCATTTTATCATTTTC 5880
QY 5881 CTGTGCTTTAAATTAACAAATGTAACAAAAAACAGATATCAAGCTGTTTAAAGTCTTCGG 5940
Db 5881 CTGTGCTTTAAATTAACAAATGTAACAAAAAACAGATATCAAGCTGTTTAAAGTCTTCGG 5940
QY 5941 CTACTGTGCCCGTGTCAATAGAGGCCCGGTTTCCCATTTGTTGACTGTGACAGGCTC 6000
Db 5941 CTACTGTGCCCGTGTCAATAGAGGCCCGGTTTCCCATTTGTTGACTGTGACAGGCTC 6000

QY 6001 AGCATGGGCTCAGACAGATGCTGCTTAATTGTGATGATACAGAAAGCCAGGCTTTGGG 6060
DB 6001 AGCATGGGCTCAGACAGATGCTGCTTAATTGTGATGATACAGAAAGCCAGGCTTTGGG 6060
QY 6061 ATACAAAGTCTTTCCTCTTCATTTGATGCCGTGCACTGTGTGAAAGAGATGTTTGTCC 6120
DB 6061 ATACAAAGTCTTTCCTCTTCATTTGATGCCGTGCACTGTGTGAAAGAGATGTTTGTCC 6120
QY 6121 GGAATATAAATATATAGTCTTGAGTCTGCCCAAAAAA 6163
DB 6121 GGAATATAAATATATAGTCTTGAGTCTGCCCAAAAAA 6163
RESULT 2
ABA9506
ID ABA9506 standard; cDNA; 5897 BP.
AC ABA9506;
XX
XX 17-MAY-2002 (first entry)
DE Human tumour-associated antigen B345 cDNA.
XX
XX Tumour-associated antigen; human; B345; cytostatic; cell communication;
KM cell interaction; signal transduction; metastasis; cancer; colon;
KM immunotherapy; carcinoma; lung; diagnosis; gene; s9.
OS Homo sapiens.
XX
XX
FH Key 1 Location/Qualifiers
FT 5'UTR 1..214
FT CDS /tag= a
FT 215..2464
FT /tag= b
FT /product= "tumour-associated antigen B345"
FT 2465..5897
FT /tag= c
XX
XX WO200204508-A1.
PN
PD 17-JAN-2002.
XX
XX
PF 05-JUL-2001; 2001WO-EP07705.
XX
XX 07-JUL-2000; 2000DE-1033080.
PR 19-APR-2001; 2001DE-1019294.
XX
XX
PA (BOEH) BOEHRINGER INGELHEIM INT GMBH.
XX
PI Schellfeller N, Scherl-Mostlgeger M, Sommergruber W, Abseher R;
XX WPI; 2002-171704/22.
DR P-PSDB; AAM49640.
XX
XX
PT New tumour-associated antigen B345, useful for diagnosis and
PT immunotherapy of tumors, also related nucleic acid and antibodies
XX
XX
PS Example 6; Page 70-76; 102pp; German.
XX
XX This invention describes a novel tumour-associated antigen, designated
CC B345 which has cytostatic activity. B345 is involved in communication,
CC interaction and/or signal transduction with extracellular components and
CC ligands, especially in the metastatic potential of cancers, particularly
CC of the colon. B345 or its immunogenic fragments, also the DNA that
CC encodes it, are useful for immunotherapy of cancer, particularly
CC carcinoma of lung or colon. Antibodies raised against B345 are useful for
CC treatment and diagnosis of cancers that are associated with B345
CC expression, including their use for targeted delivery of cytotoxic or
CC radioactive agents. Probes derived from B345 can be used to detect
CC tumour-specific mutations in the B345 sequence, and can be used to screen
CC for B345 specific modulators. This sequence encodes a human B345
CC tumour-associated antigen described in the invention.

XX
SQ Sequence 5897 BP; 1464 A; 1596 C; 1364 G; 1467 T; 6 other:
Query Match 95.0%; Score 5854.2; DB 24; Length 5897;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 5865; Conservative 2; Mismatches 5; Indels 1; Gaps 1;
QY 291 CCTGAATCGGGGCTCTATCGCACTGCTAGAGGGTTCGTCGCGGGGCGGCGCC 350
DB 25 CCGAACTCGGGGCTCTATCGCACTGCTAGAGGGTTCGTCGCGGGGCGGCGCC 84
QY 351 GCCCGCGGGGCGAAGCTTTGAGATTCGTCGCCACGAAAGCAACTTACATTC 410
DB 85 GCCCGCGGGGCGAAGCTTTGAGATTCGTCGCCACGAAAGCAACTTACATTC 144
QY 411 CATAAAGCTGGGACCCCGACTGCTGCGCAAAACCTGTTACATGTCATTTCTAAAG 470
DB 145 CATAAAGCTGGGACCCCGACTGCTGCGCAAAACCTGTTACATGTCATTTCTAAAG 204
QY 471 ACATATACCATGTTGTCCATCAAGTCTGGAGAAAGATATGCTTTACCTTAGTCCCA 530
DB 205 ACATATACCATGTTGTCCATCAAGTCTGGAGAAAGATATGCTTTACCTTAGTCCCA 264
QY 531 GAGTCTGAGAAATCACCTTTGTCAATAGAGATCCAGAAATATGACTGTATGACAGCC 590
DB 265 GAGTCTGAGAAATCACCTTTGTCAATAGAGATCCAGAAATATGACTGTATGACAGCC 324
QY 591 ATGTCTTTTGGGAGGTTCAAGCTTCAGCCCTGACATCGTTGTTGCTTACCTCAACAG 650
DB 325 ATGTCTTTTGGGAGGTTCAAGCTTCAGCCCTGACATCGTTGTTGCTTACCTCAACAG 384
QY 651 AACTTTCATCTGGGATGTCAAAGCTCATAGACATTCGTTTAGAGTGTGAGTTTCCAT 710
DB 385 AACTTTCATCTGGGATGTCAAAGCTCATAGACATTCGTTTAGAGTGTGAGTTTCCAT 444
QY 711 CCTCGCTGAGCAATCGTCCGGGTGAGAGCTGCCAGAGAGTCACTATCCAT 770
DB 445 CCTCGCTGAGCAATCGTCCGGGTGAGAGCTGCCAGAGAGTCACTATCCAT 504
QY 771 CAGCGCGGAAATGATGATGCCAGCTGTGAGAGTGGAACTTGTGACGAATGGCACTGT 830
DB 505 CAGCGCGGAAATGATGATGCCAGCTGTGAGAGTGGAACTTGTGACGAATGGCACTGT 564
QY 831 GTCCCGGAAATGATGATGCCAGCTGTGAGAGTGGAACTTGTGACGAATGGCACTGT 890
DB 565 GTCCCGGAAATGATGATGCCAGCTGTGAGAGTGGAACTTGTGACGAATGGCACTGT 624
QY 891 CAGAAATGTCCTCGGCTTCAGATTGCAACCGCTCATATATAAGCTGTGATCAT 950
DB 625 CAGAAATGTCCTCGGCTTCAGATTGCAACCGCTCATATATAAGCTGTGATCAT 684
QY 951 CGAGTCTGTTGAGGGTGAAGGCTCAGCAACCTGATGTGCAACTACCAAGAG 1010
DB 685 CGAGTCTGTTGAGGGTGAAGGCTCAGCAACCTGATGTGCAACTACCAAGAG 744
QY 1011 CTTCCTGAGGATGAGTCTATGACGTGGCAGTTTGTCTCTGCGACACTGCGGGCCAG 1070
DB 745 CTTCCTGAGGATGAGTCTATGACGTGGCAGTTTGTCTCTGCGACACTGCGGGCCAG 804
QY 1071 CGTCTCTTCTCAACTTCAACTTCCAACTGTGAGAGAGAGAGAGGAGGTTGAAT 1130
DB 805 CGTCTCTTCTCAACTTCAACTTCCAACTGTGAGAGAGAGAGAGGAGGTTGAAT 864
QY 1131 CTACATCCGCGGCTCAGCAACCAACCCGAGTGTTCAACTGGAGAGCAAGAGCTGG 1190
DB 865 CTACATCCGCGGCTCAGCAACCAACCCGAGTGTTCAACTGGAGAGCAAGAGCTGG 924
QY 1191 GAACATGGGCGGGAATTCACCTCTCTGCAAGGCTGTGACCAAGATCCCAAGTCC 1250
DB 925 GAACATGGGCGGGAATTCACCTCTCTGCAAGGCTGTGACCAAGATCCCAAGTCC 984
QY 1251 AGGATCTTCGCGCTCAGTTCAGTTTGGTCCAAATCCCAAAATGAAGCAATAA 1310
DB 984 AGGATCTTCGCGCTCAGTTCAGTTTGGTCCAAATCCCAAAATGAAGCAATAA 1310

Db 985 AGGATCCTCCGGCTGAGTTCCAAAGTTTGGTCCAAACATCCACAAATGAAACAAATA 1044
QY 1311 AATTCAGCTGTTACTTGAATAATGAGCGAGCATGTCACTCAACATCGACGACGCGC 1370
Db 1045 AATTCAGCTGTTACTTGAATAATGAGCGAGCATGTCACTCAACATCGACGACGCGC 1104
QY 1371 CGTCAACAGAGCGGCAAGTTGTCCTGGCTGTTGGTGTGTAGAAATCTCGGACTG 1430
Db 1105 CGTCAACAGAGCGGCAAGTTGTCCTGGCTGTTGGTGTGTAGAAATCTCGGACTG 1164
QY 1431 CAGTAGCAACCTCAACCTGACATCTGGCTCCAAACACAAATCTCTTCTTGTGATGA 1490
Db 1165 CAGTAGCAACCTCAACCTGACATCTGGCTCCAAACACAAATCTCTTCTTGTGATGA 1224
QY 1491 TCTGACACGCTGTGTGATGAATGTGAAAAAACATAAGCTGACAGACACCGGTA 1550
Db 1225 TCTGACACGCTGTGTGATGAATGTGAAAAAACATAAGCTGACAGACACCGGTA 1284
QY 1551 CCAAGAGAAATCTTACCTACCTCCAGGTCAGAGTACATCTCCACCTGCTGTGAGACT 1610
Db 1285 CCAAGAGAAATCTTACCTACCTCCAGGTCAGAGTACATCTCCACCTGCTGTGAGACT 1344
QY 1611 GCATGACTTCTCTGGAAGCTGTGTGTGCCAAGAGACAGGCTCAAGCTGGTGTGTGCC 1670
Db 1345 GCATGACTTCTCTGGAAGCTGTGTGTGCCAAGAGACAGGCTCAAGCTGGTGTGTGCC 1404
QY 1671 AGCCCAAGCTGACAGACATACACAGAGAACCTTCGAACACAGCTTCAGCTACT 1730
Db 1405 AGCCCAAGCTGACAGACATACACAGAGAACCTTCGAACACAGCTTCAGCTACT 1464
QY 1731 CGTGGCAGAGTCCATACCCAGGTCAGGACCTGTACTTGGCTCCTTGTGCCGGAGGCTC 1790
Db 1465 CGTGGCAGAGTCCATACCCAGGTCAGGACCTGTACTTGGCTCCTTGTGCCGGAGGCTC 1524
QY 1791 TATCAACAGATCCAGGTGAAGAGCAACATCTCGGTGACCTTGGACCTTGTGCCCGAG 1850
Db 1525 TATCAACAGATCCAGGTGAAGAGCAACATCTCGGTGACCTTGGACCTTGTGCCCGAG 1584
QY 1851 CTTTCAACAGAGGCTTCACAGGACAGGCTGTGACGCTGCTTTTATTTTCAAGA 1910
Db 1585 CTTTCAACAGAGGCTTCACAGGACAGGCTGTGACGCTGCTTTTATTTTCAAGA 1644
QY 1911 GGAAGGGGTTTTCAGGGTGAACCCCTGACACAAAAAGCAAGGTCTACCTGAGGACCCCAA 1970
Db 1645 GGAAGGGGTTTTCAGGGTGAACCCCTGACACAAAAAGCAAGGTCTACCTGAGGACCCCAA 1704
QY 1971 CTGGGACCGGGGCTGACATCTCTGACCTGTGTGTCTCTGGAACATCAGCGTGCCAGAGA 2030
Db 1705 CTGGGACCGGGGCTGACATCTCTGACCTGTGTGTCTCTGGAACATCAGCGTGCCAGAGA 1764
QY 2031 CCAGGTGGCTGCTGACATCTCTTAAAGAGCGGAGCGGCTGTGCGACAGAGGCG 2090
Db 1765 CCAGGTGGCTGCTGACATCTCTTAAAGAGCGGAGCGGCTGTGCGACAGAGGCG 1824
QY 2091 CGCATTCATGATCCAGAGGAGCGGAGCCGGGCTGAGGAGTCTTCAAGCTGAGAGA 2150
Db 1825 CGCATTCATGATCCAGAGGAGCGGAGCCGGGCTGAGGAGTCTTCAAGCTGAGAGA 1884
QY 2151 GGATGTCTCCCAAGCAAGCTTCACATCAAGCTTGTGGGTCAACAATCTTAACTG 2210
Db 1885 GGATGTCTCCCAAGCAAGCTTCACATCAAGCTTGTGGGTCAACAATCTTAACTG 1944
QY 2211 CAGCCCAAGAGCGGAGCAAGCTAGACGCTGCTTCTGCGTGACACTTACCCCAAGAC 2270
Db 1945 CAGCCCAAGAGCGGAGCAAGCTAGACGCTGCTTCTGCGTGACACTTACCCCAAGAC 2004
QY 2271 TGTGAGCTTGAATGATCTCTCATCGAGCGGAGGAGGTGTGAGTCTTACTGCTGTGC 2330
Db 2005 TGTGAGCTTGAATGATCTCTCATCGAGCGGAGGAGGTGTGAGTCTTACTGCTGTGC 2064
QY 2331 CCTCGGCTCATCTTGTGTGTGAAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAG 2390
Db 2065 CCTCGGCTCATCTTGTGTGTGTGAAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAG 2124

QY 2391 TGTGCTATCTACATGAGCAACATCAATCTGAGATGCCGAGGACGCCAAAAAGTTTCA 2450
Db 2125 TGTGCTATCTACATGAGCAACATCAATCTGAGATGCC-AGGACGCCAAAAAGTTTCA 2183
QY 2451 GAAAGGGGAAAGGACATGATGCTCCATGTGTATGAGCTCATGAGGACACCATGTGTTA 2510
Db 2184 GAAAGGGGAAAGGACATGATGCTCCATGTGTATGAGCTCATGAGGACACCATGTGTTA 2243
QY 2511 TGGCATCTGCTACAGATTCACAGCGGCTCCTTCTGTGAGGACAGAGTGAACACTACG 2570
Db 2244 TGGCATCTGCTACAGATTCACAGCGGCTCCTTCTGTGAGGACAGAGTGAACACTACG 2303
QY 2571 GCCGTTCCAGGGACACATGGGGGTGTCTCTCCCTCCCAACCATATGCTCCAGGGC 2630
Db 2304 GCCGTTCCAGGGACACATGGGGGTGTCTCTCCCTCCCAACCATATGCTCCAGGGC 2363
QY 2631 CCCAAGTGAAGTTGGCCACTGAGGAGCCACTCTCTGCTCCCTCTGTAGTGTGAGAG 2690
Db 2364 CCCAAGTGAAGTTGGCCACTGAGGAGCCACTCTCTGCTCCCTCTGTAGTGTGAGAG 2423
QY 2691 TGAACGCTACACTTCTCCCATCCCAACAAATGGGATGTAAAGAGCAAGAGACAGACAT 2750
Db 2424 TGAACGCTACACTTCTCCCATCCCAACAAATGGGATGTAAAGAGCAAGAGACAGACAT 2483
QY 2751 TCCCTTACTGAACACTGAGAGCCATGAGACAGAGAAATTAATGATCCATTCAGAC 2810
Db 2484 TCCCTTACTGAACACTGAGAGCCATGAGACAGAGAAATTAATGATCCATTCAGAC 2543
QY 2811 GCTTGTGATGTTTCAATGAAGAGGAGGACAGAGACACCCGCTGTTTCTTAACAGAAA 2870
Db 2544 GCTTGTGATGTTTCAATGAAGAGGAGGACAGAGACACCCGCTGTTTCTTAACAGAAA 2603
QY 2871 TCCTTAAGAGAGGAAATTAACAGAAAGAGAGAGGTTTCTGAGACCCGCAAC 2930
Db 2604 TCCTTAAGAGAGGAAATTAACAGAAAGAGAGAGGTTTCTGAGACCCGCAAC 2663
QY 2931 TTCATATTGCTCAGTGACATCTTAAGGCAAGACATTTGAAAAATGATTAATTCACATC 2990
Db 2664 TTCATATTGCTCAGTGACATCTTAAGGCAAGACATTTGAAAAATGATTAATTCACATC 2723
QY 2991 TGGATACAGTATGACAGCTGATGCTGCTCAACTTAAGGCTGAGGTTTGGACAGCTG 3050
Db 2724 TGGATACAGTATGACAGCTGATGCTGCTCAACTTAAGGCTGAGGTTTGGACAGCTG 2783
QY 3051 TTAATGAGAGGAGAGGCTGAGTCACTAGCATAGGTTTGCACAGACCTTGATTCAG 3110
Db 2784 TTAATGAGAGGAGAGGCTGAGTCACTAGCATAGGTTTGCACAGACCTTGATTCAG 2843
QY 3111 AGTGTAAACAGAGGCTTGGCTTCTTCAAGACAACTTCCAAATTCGAAGAGCTTACT 3170
Db 2844 AGTGTAAACAGAGGCTTGGCTTCTTCAAGACAACTTCCAAATTCGAAGAGCTTACT 2903
QY 3171 GAGTCCCTACTCTACATGGGGTCCCGAGTGAAGAACGAATGTGCTTTTATATT 3230
Db 2904 GAGTCCCTACTCTACATGGGGTCCCGAGTGAAGAACGAATGTGCTTTTATATT 2963
QY 3231 ATTTATTTGGTGTCTGTGTATTTAAAGATCAAAATGTATAACCACTAGCTCTTTTC 3290
Db 2964 ATTTATTTGGTGTCTGTGTATTTAAAGATCAAAATGTATAACCACTAGCTCTTTTC 3023
QY 3291 ACCTGACTTGAATTAATCACTAATCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3350
Db 3024 ACCTGACTTGAATTAATCACTAATCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3083
QY 3351 ACCGCTAGTAAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3410
Db 3084 ACCGCTAGTAAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3143
QY 3411 CCAAGAAAGATGTGTGTGTGAGAGCAATTCACATATCTGCTTGTGATTAAGAGACTT 3470
Db 3144 CCAAGAAAGATGTGTGTGTGAGAGCAATTCACATATCTGCTTGTGATTAAGAGACTT 3203

3471 CCTGATTCCTAGAGCGGTCTGTTATCCATGCGAAATTCATCTGAATCCCAT 3530
3204 CCGATTCCTAGAGCGGTCTGTTATCCATGCGAAATTCATCTGAATCCCAT 3263
3531 GTCCTATAGCTAGCAATAGAGAAATTCCTCAAGTTTCATGTCGGTTCTCTAGC 3590
3264 GTCCTATAGCTAGCAATAGAGAAATTCCTCAAGTTTCATGTCGGTTCTCTAGC 3323
3591 TCGAGCAATACCTTGGACATTTAAAGAAATTTAGAGAAATTTCTCTCTCTAAATG 3650
3324 TCGAGCAATACCTTGGACATTTAAAGAAATTTAGAGAAATTTCTCTCTCTAAATG 3383
3651 TTTAAATATATACCAACAGTGGCCCTCATAGTTTCTGTGGCACTGCAACCCAT 3710
3384 TTTAAATATATACCAACAGTGGCCCTCATAGTTTCTGTGGCACTGCAACCCAT 3443
3711 TACTTGTAGCTTAAAAACAACATTAAGTTATAGTCCTGAGGATCAGAAATTC 3770
3444 TACTTGTAGCTTAAAAACAACATTAAGTTATAGTCCTGAGGATCAGAAATTC 3503
3771 GGTATGCCCTGATGAAATTCAGAGTGTACAGAGCTGTCTCTCTGAGAGCTTAG 3830
3504 GGTATGCCCTGATGAAATTCAGAGTGTACAGAGCTGTCTCTCTGAGAGCTTAG 3563
3831 GGAGAAAGCGGTCTCTGCAATTTCAAGCTTCTAGAGGCTGGGTCATTCGAGGCTCA 3890
3564 GGAGAAAGCGGTCTCTGCAATTTCAAGCTTCTAGAGGCTGGGTCATTCGAGGCTCA 3623
3891 GTGGCTGTGCAAGCTTTTCTACATGGCATCACTGTGACACTGGCCCTCCACCTTCC 3950
3624 GTGGCTGTGCAAGCTTTTCTACATGGCATCACTGTGACACTGGCCCTCCACCTTCC 3683
3951 TTTGACTTACAAAGCCCAACAGAGATCCAGATATCTCTCATCTAAAGATCTTCA 4010
3684 TTTGACTTACAAAGCCCAACAGAGATCCAGATATCTCTCATCTAAAGATCTTCA 3743
4011 TCATCTGGAAGAGCTTTTGGCATGCAAGACAAATAGGCACAGGGGGGATTAGAGAC 4070
3744 TCATCTGGAAGAGCTTTTGGCATGCAAGACAAATAGGCACAGGGGGGATTAGAGAC 3803
4071 AGGACATCTTTGGGCTGCTTATTTCTGCTACACACCTTCTGCTGACATGATCCCA 4130
3804 AGGACATCTTTGGGCTGCTTATTTCTGCTACACACCTTCTGCTGACATGATCCCA 3863
4131 GGAGAGCTTCAAAATATGATGTGGCGACAGAGATTTTGTTTTACCTTGGAGCTTAC 4190
3864 GGAGAGCTTCAAAATATGATGTGGCGACAGAGATTTTGTTTTACCTTGGAGCTTAC 3923
4191 ACTTAAAAAACCAGATCAGAAAGATCTGGSCATGCTGGGGTCATCTCACTAGC 4250
3924 ACTTAAAAAACCAGATCAGAAAGATCTGGSCATGCTGGGGTCATCTCACTAGC 3983
4251 AACAGCTGCTGAGCTGGGACACAGCTGTGCTTTAGAGGGGTGTCTACTCACAGG 4310
3984 AACAGCTGCTGAGCTGGGACACAGCTGTGCTTTAGAGGGGTGTCTACTCACAGG 4043
4311 TCACACAGCCACACACTAGCCCTATCTATCTCCACATAGAGCTAAGTGTGTCTA 4370
4044 TCACACAGCCACACACTAGCCCTATCTATCTCCACATAGAGCTAAGTGTGTCTA 4103
4371 CTATCAATGCGCCTGAGGTGCTATTTATTTAAATGAAAGAAACAGTGGATTAATC 4430
4104 CTATCAATGCGCCTGAGGTGCTATTTATTTAAATGAAAGAAACAGTGGATTAATC 4163
4431 TCTAATCAGGTAGTAGACCATAGACCAATGTGTCTCATATTACCTTTTCTTTT 4490
4164 TCTAATCAGGTAGTAGACCATAGACCAATGTGTCTCATATTACCTTTTCTTTT 4223
4491 TTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 4550
4224 TTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 4283
4551 TGGAGTGCATGTGGGCAATCTGGCTCAGTGCACCTTCTGCTTCTGGGCTCAAGCAAT 4610
4284 TGGAGTGCATGTGGGCAATCTGGCTCAGTGCACCTTCTGCTTCTGGGCTCAAGCAAT 4343
4611 CTCCTACCTAGCCTCCCAATAGCTGGATCACTGGCACAAACACCATAGCCAGCTAA 4670
4344 CTCCTACCTAGCCTCCCAATAGCTGGATCACTGGCACAAACACCATAGCCAGCTAA 4403
4671 TTTTGTATTTTGTATAGACAGGTTTCAACAGTTTCCCAAGCTGTCTCAACCTTCT 4730
4404 TTTTGTATTTTGTATAGACAGGTTTCAACAGTTTCCCAAGCTGTCTCAACCTTCT 4463
4731 GGGCTCAAGCAATCTCTGCTGGCTCCCAAGTGTGGATTTACAGATGAGACCA 4790
4464 GGGCTCAAGCAATCTCTGCTGGCTCCCAAGTGTGGATTTACAGATGAGACCA 4523
4791 CCGCATCAGCCCCACACCTCATTTATACCAATTTACCTGGCCAGTAACTGTGGACTTT 4850
4524 CCGCATCAGCCCCACACCTCATTTATACCAATTTACCTGGCCAGTAACTGTGGACTTT 4583
4851 GCTTCCCTACCCCTGCTGTGATCTGGAAGAGAGGATTTATGTTATAGCTTTCAGACA 4910
4584 GCTTCCCTACCCCTGCTGTGATCTGGAAGAGAGGATTTATGTTATAGCTTTCAGACA 4643
4911 GTCCCAAGTTCAATATTTCTGCGGCAAAAACCTTCTCAAAAAATTAATGACTTCAAT 4970
4644 GTCCCAAGTTCAATATTTCTGCGGCAAAAACCTTCTCAAAAAATTAATGACTTCAAT 4703
4971 TATTCATGAAATTCACCTTGGAAATGACCGCTCAACTGTTTCACATGCAATTAATGAA 5030
4704 TATTCATGAAATTCACCTTGGAAATGACCGCTCAACTGTTTCACATGCAATTAATGAA 4763
5031 AGGAATTTTATAGTCTCTTAAATGGCGTAACTGCAAGCTCTGTAACCTTCCAGAG 5090
4764 AGGAATTTTATAGTCTCTTAAATGGCGTAACTGCAAGCTCTGTAACCTTCCAGAG 4823
5091 GATAGATATTTAAGTATGCTATGCTTGGCGTGTGCTTAAAGGACCTTCTCTGAAAGTC 5150
4824 GATAGATATTTAAGTATGCTATGCTTGGCGTGTGCTTAAAGGACCTTCTCTGAAAGTC 4883
5151 TGGTCTGCGCCAGTGAACCTTGGCTTGTGAGGCGGAGATGACCTGATTAAGGGC 5210
4884 TGGTCTGCGCCAGTGAACCTTGGCTTGTGAGGCGGAGATGACCTGATTAAGGGC 4943
5211 CAAGAGAGGCTGTGGCTCTCTTCCCTCAGTGAAGAGCCCTTATTTGAATTCAGTGTG 5270
4944 CAAGAGAGGCTGTGGCTCTCTTCCCTCAGTGAAGAGCCCTTATTTGAATTCAGTGTG 5003
5271 GAGCCCTAGCCCTCATTTCTGACATTTCCCAACCTCCAGCCCTTCCAAAGAGACTA 5330
5004 GAGCCCTAGCCCTCATTTCTGACATTTCCCAACCTCCAGCCCTTCCAAAGAGACTA 5063
5331 GGTGCTGCTGATTCACCAAGTGTGGATTTGGCTTCTTAAGCTGACTTGTCAACA 5390
5064 GGTGCTGCTGATTCACCAAGTGTGGATTTGGCTTCTTAAGCTGACTTGTCAACA 5123
5391 TCACCGCATCATCTGTGCTGTGCAAGACACACGTGGCAATTTTCTTCAACTGAGAGG 5450
5124 TCACCGCATCATCTGTGCTGTGCAAGACACACGTGGCAATTTTCTTCAACTGAGAGG 5183
5451 TCAAACTCTCGGCAAGTGTGGCTGTGAGACACAGTATTTCTGAGAGTGGCTCA 5510
5184 TCAAACTCTCGGCAAGTGTGGCTGTGAGACACAGTATTTCTGAGAGTGGCTCA 5243
5511 GTGAAGGGGCCACCTGAGGAACCTGTGCTTTTCTTTAAAGCCAGGCCCACTTAC 5570
5244 GTGAAGGGGCCACCTGAGGAACCTGTGCTTTTCTTTAAAGCCAGGCCCACTTAC 5303
5571 ATTAACATTTTCAAGGCTCACTGGAAACAGTAAATGCCATTTGTGAAGCTTACTGATG 5630
5304 ATTAACATTTTCAAGGCTCACTGGAAACAGTAAATGCCATTTGTGAAGCTTACTGATG 5370
5631 CCAAGCCAGTCTCATCAGTGGTGTGCAATGCTGCTGAGGAAGGCGAGGCGACTG 5690

Db 5364 CCACCCACCTCTATCCACCTGCTATGCCATGACAGAAAGCCAGCCATGACGG 5423
 Qy 5691 ACTGCTCTATATGCTGTGCTATGTCACAGAGAGGAAAGCTCTCAAGAGAGTCACT 5750
 Db 5424 ANTGGTCTTAATGNTGNTGGTCAATGACAGAGAAAGGTCTCAAGAGAGTCACT 5483
 Qy 5751 GGGACAGACAGACAGCCACCGGACATGGCTTGTAAAGTTAGACAGACTGCTGTGTG 5810
 Db 5484 GGGACAGACAGACAGCCACCGGACATGGCTTGTAAAGTTAGACAGACTGCTGTGTG 5543
 Qy 5811 GATCTGACAGTCTCTACAGTGAATATATTTATTCATTCAGATACCTTTTAGTGGCATTT 5870
 Db 5544 GATCTGACAGTCTCTACAGTGAATATATTTATTCATTCAGATACCTTTTAGTGGCATTT 5603
 Qy 5871 TATTCATTTCTCTGCTTTAAATGAACAAATGTACCAAAAACAAGATCAAGCTGTCTTA 5930
 Db 5604 TATTCATTTCTCTGCTTTAAATGAACAAATGTACCAAAAACAAGATCAAGCTGTCTTA 5663
 Qy 5931 AGTCTTCGCTACTTGTGCTCCCTGGTTCAGTAGAGCCCGGTTCCAGTTGTGACTG 5990
 Db 5664 AGTCTTCGCTACTTGTGCTCCCTGGTTCAGTAGAGCCCGGTTCCAGTTGTGACTG 5723
 Qy 5991 TGACAGGCTCAGATGGGCTCAGCAGATGCTGCTTAATTTGTGATGATACAGAAAGCC 6050
 Db 5724 TGACAGGCTCAGATGGGCTCAGCAGATGCTGCTTAATTTGTGATGATACAGAAAGCC 5783
 Qy 6051 AGCTTTGGATACAGATTTCTTCTCTTCAATTTGATGCCGTGACATGTGTGAAGCAGAT 6110
 Db 5784 AGCTTTGGATACAGATTTCTTCTCTTCAATTTGATGCCGTGACATGTGTGAAGCAGAT 5843
 Qy 6111 GTTTTGTCCGGAATAAATAATATAGTCTGTGAGTCTCCGCCAAAAA 6163
 Db 5844 GTTTTGTCCGGAATAAATAATATAGTCTGTGAGTCTCCGCCAAAAA 5896

RESULT 3

AAH18243
ID AAH18243 standard; cDNA: 5573 BP.

AC AAH18243:

DT 26-JUN-2001 (first entry)

DE Human cDNA sequence SEQ ID NO:18190.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

OS Homo sapiens.

PN EP1074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99DP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

PI Oka T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -

PS Claim 8; SEQ ID 18190; 2537bp + CD ROM; English.
 XX The present invention describes primer sets for syn-
 CC full-length cDNAs defined in the specification. The
 CC comprises: (a) an oligo-dT primer and an oligonuc-
 CC to the complementary strand of a polynucleotide whi-
 CC the 5602 nucleotide sequences defined in the specifi-
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the combina-
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX Sequence 5573 BP; 1373 A; 1524 C; 1290 G; 1386 T; 0 other;

Query Match 89.5%; Score 5518.4; DB 22; Length 5573;

Best Local Similarity 99.8%; Pred. No. 0; Matches 5566; Conservative 0; Mismatches 6; Indels 4; Gaps 4;

Qy 569 ATATTGACTGATGTCAGAGCCCATGTCCTTTTGGGAGGTTCAAGCTTCAGCTGACAT 628
 Db 1 ATATTGACTGATGTCAGAGCCCATGTCCTTTTGGGAGGTTCAAGCTTCAGCTGACAT 60
 Qy 629 CGTGTGCCATCCCTCAACGAACTTTCATCTGGAATGTCATTAAGCTCATTAAGCAATCG 688
 Db 61 CGTGTGCCATCCCTCAACGAACTTTCATCTGGAATGTCATTAAGCTCATTAAGCAATCG 120
 Qy 689 GTTTAGAGCTCAGATTTTCCATCCCTGCGCTGAGAGAGATGTCGCGGTGAGAGCTGCC 748
 Db 121 G-TTAGAGCTCAGATTTTCCATCCCTGCGCTGAGAGAGATGTCGCGGTGAGAGCTGCC 179
 Qy 749 CAGACGAGTACATCTCATCAGCGCGCAATGATGCCACCGTGTGATCGGAA 808
 Db 180 CAGACGAGTACATCTCATCAGCGCGCAATGATGCCACCGTGTGATCGGAA 239
 Qy 809 CCTTCTGAGCAATGGGACATGTCGCGGATCAAGATGCAAGAGAGTGAATGGCT 868
 Db 240 CCTTCTGAGCAATGGGACATGTCGCGGATCAAGATGCAAGAGAGTGAATGGCT 299
 Qy 869 TACACCTCCATGCTTCCACCCCAAGAAATGTCGCGGCTTCCAGCATTCGCAACCGCTCAT 928
 Db 300 TACACCTCCATGCTTCCACCCCAAGAAATGTCGCGGCTTCCAGCATTCGCAACCGCTCAT 359
 Qy 929 CTATTAACAGCTCTGTCATTCAGAGTGTGTTGAGGTTGAAGGCTCAGCAACCTCGA 988
 Db 360 CTATTAACAGCTCTGTCATTCAGAGTGTGTTGAGGTTGAAGGCTCAGCAACCTCGA 419
 Qy 989 TGCTGCAACTACCAAGAGGCTTCCCTGAGATGAGCTCATACGCGGAGTTGTGCG 1048
 Db 420 TGCTGCAACTACCAAGAGGCTTCCCTGAGATGAGCTCATACGCGGAGTTGTGCG 479
 Qy 1049 TTCTGCAACCTGCGGAGGCTGCTCTCTCTCTCAACTTCACCTTCACAACTGTGAGA 1108
 Db 480 TTCTGCAACCTGCGGAGGCTGCTCTCTCTCTCTCAACTTCACCTTCACAACTGTGAGA 539
 Qy 1109 GGAAGAGAGAGCGGTTGAATATCAATCCCGGCTCCACCAACCCGAGGTGTCA 1168
 Db 540 GGAAGAGAGAGCGGTTGAATATCAATCCCGGCTCCACCAACCCGAGGTGTCA 599
 Qy 1169 AGCTGAGAGACACACCTGCGGAACATGGCGGGAATTAACCTCTCTCTGCAAGCT 1228

Dh 600 AGCTGGAGGACAGACAGCTGGGAAACATGCGGGGAACCTTCAACCTCTCTGCAAGGCT 659
Qy 1229 GTGACCAAGATGCGCCAAAGATCCAGGGAATCTCCGGCTGAGTTCCAAAGTTTGGTCCAC 1288
Dh 660 GTGACCAAGATGCGCCAAAGATCCAGGGAATCTCCGGCTGAGTTCCAAAGTTTGGTCCAC 719
Qy 1289 ATCCACAAAATGAAGAAATCAAAATCTAGCTGGTGGTGGTAAATGAGCAGGACATGT 1348
Dh 720 ATCCACAAAATGAAGAAATCAAAATCTAGCTGGTGGTGGTAAATGAGCAGGACATGT 779
Qy 1349 CACTACCATCGAGCCAGCCCGCTCAAAACAGAGCCGCAAGTTTGTCCCTGGCTGTTTCG 1408
Dh 780 CACTACCATCGAGCCAGCCCGCTCAAAACAGAGCCGCAAGTTTGTCCCTGGCTGTTTCG 839
Qy 1409 TGTGTCTAATCTGGGACCTGAGTGAAGAACCTGACCTGACATTTGGCTCCAAACACA 1468
Dh 840 TGTGTCTAATCTGGGACCTGAGTGAAGAACCTGACCTGACATTTGGCTCCAAACACA 899
Qy 1469 AATCTCTCTCTCTTGTGATCTGACACGTTGTGATGAAATGTGAAAAACCATTA 1528
Dh 900 AATCTCTCTCTCTTGTGATCTGACACGTTGTGATGAAATGTGAAAAACCATTA 959
Qy 1529 GCTGCACAGACACCGGTAAGTCCGCAAAAGAAATCTACTACTCCAGGTGCCAGTGA 1588
Dh 960 GCTGCACAGACACCGGTAAGTCCGCAAAAGAAATCTACTACTCCAGGTGCCAGTGA 1019
Qy 1589 TCTCTCACTGGCTGTGAGCTGATGACTTCTCTGGAAGCTGTGTGTTCCCAAGACA 1648
Dh 1020 TCTCTCACTGGCTGTGAGCTGATGACTTCTCTGGAAGCTGTGTGTTCCCAAGACA 1079
Qy 1649 GGCTCAGCTGTGTGTGTGAGCCAGCCAGAGCTGACAGATACACAGAGAACCCCT 1708
Dh 1080 GGCTCAGCTGTGTGTGTGAGCCAGCCAGAGCTGACAGATACACAGAGAACCCCT 1139
Qy 1709 GCAACACAGCTTCACTACTCTGTGCGCAAGTCCATACCAGCAGAGACTGTACTTGC 1768
Dh 1140 GCAACACAGCTTCACTACTCTGTGCGCAAGTCCATACCAGCAGAGACTGTACTTGC 1199
Qy 1769 GCTCTTCTGCGCGGAGGCTCTATCAGACATCCAGGTGAAGCAGAAATCTCGGTGA 1828
Dh 1200 GCTCTTCTGCGCGGAGGCTCTATCAGACATCCAGGTGAAGCAGAAATCTCGGTGA 1259
Qy 1829 CCCTTGCACCTTTGCCCCAGCTTCCAAAGAGGCTCCAGGCGAGGGCTGACGGTGT 1888
Dh 1260 CCCTTGCACCTTTGCCCCAGCTTCCAAAGAGGCTCCAGGCGAGGGCTGACGGTGT 1319
Qy 1889 CCTTATACCTTATTTCAAGAGAGAGGCTTTTCAAGGTGACCCCTGACACAAAAAGCA 1948
Dh 1320 CCTTATACCTTATTTCAAGAGAGAGGCTTTTCAAGGTGACCCCTGACACAAAAAGCA 1379
Qy 1949 AGCTTACCTGAGAGACCCCAACTGGGACCGGGGCTGCCATCCCTGACCTGTGTCT 2008
Dh 1380 AGCTTACCTGAGAGACCCCAACTGGGACCGGGGCTGCCATCCCTGACCTGTGTCT 1439
Qy 2009 GGAACATCAGCGGTGCCAGAGACAGGTGGCTGGCTGACTTCTTTAAGAGAGGAGG 2068
Dh 1440 GGAACATCAGCGGTGCCAGAGACAGGTGGCTGGCTGACTTCTTTAAGAGAGGAGG 1499
Qy 2069 GCGTGGTGTCCAGAGAGGCGCGCTTATGATCATTCAGAGAGCAGCGAGCCGGGCTG 2128
Dh 1500 GCGTGGTGTCCAGAGAGGCGCGCTTATGATCATTCAGAGAGCAGCGAGCCGGGCTG 1559
Qy 2129 AGGAGATCTTCAAGCTTGAAGAGATGTCTCCCAAGCCAAAGCTTCCACATCAGCT 2188
Dh 1560 AGGAGATCTTCAAGCTTGAAGAGATGTCTCCCAAGCCAAAGCTTCCACATCAGCT 1619
Qy 2189 TCTGGGTCAACATCTTAATCTGAGCCCAAGAGGCGGCAAGAGCTAGAGCTGTCTCT 2248
Dh 1620 TCTGGGTCAACATCTTAATCTGAGCCCAAGAGGCGGCAAGAGCTAGAGCTGTCTCT 1679
Qy 2249 CGGTGACACTTACCCCAAGAGAGTGTGACTGTGACTGTCACTCCATCGCAGCGGCTGGAG 2308
Dh 1680 CGGTGACACTTACCCCAAGAGAGTGTGACTGTGACTGTCACTCCATCGCAGCGGCTGGAG 1739

Qy 2309 GTGAGAGCTTACTGCTGTGCGCCCTGCGGCTCATCTTGGCTGTGAAAAAGAAAA 2368
Dh 1740 GTGAGAGCTTACTGCTGTGCGCCCTCGGCTCATCTTGGCTGTGAAAAAGAAAA 1799
Qy 2369 AGAAGACAAACAAAGGCGCCGCTGTGATCTTCAATGAGCAATCAATCTAGATGC 2428
Dh 1800 AGAAGACAAACAAAGGCGCCGCTGTGATCTTCAATGAGCAATCAATCTAGATGC 1859
Qy 2429 CGAGGACCCCAAAAAGTTTCAGAAAAGGCGAAAGCAATGACTCCATGTGTATGAC 2488
Dh 1860 CGAGGACCCCAAAAAGTTTCAGAAAAGGCGAAAGCAATGACTCCATGTGTATGAC 1919
Qy 2489 TCAATGAGGACACATGATGTATGGGATCTGTACAGGATTCAGGGGCTCTCTCTGC 2548
Dh 1920 TCAATGAGGACACATGATGTATGGGATCTGTACAGGATTCAGGGGCTCTCTCTGC 1979
Qy 2549 AGCCAGAGGTGACACCTTACCGGCGCTTCCAGGCGACCATGGGGCTGTCTCCCTCC 2608
Dh 1980 AGCCAGAGGTGACACCTTACCGGCGCTTCCAGGCGACCATGGGGCTGTCTCCCTCC 2039
Qy 2609 CACCCACATATGCTCCAGGCGCCCAACTGCAAAAGTTGGCCACTGAGAGCCACTCTC 2668
Dh 2040 CACCCACATATGCTCCAGGCGCCCAACTGCAAAAGTTGGCCACTGAGAGCCACTCTC 2099
Qy 2669 GCTCCCTCTCTGAGTGTGAGAGTGAACCCGTACACCTTCTCCATCCCAATAGGGGATG 2728
Dh 2100 GCTCCCTCTCTGAGTGTGAGAGTGAACCCGTACACCTTCTCCATCCCAATAGGGGATG 2159
Qy 2729 TAAGCAGCAAGAGACAGACATTCCTTACTGAACACTAGAGGCCATGAGCCAGAC 2788
Dh 2160 TAAGCAGCAAGAGACAGACATTCCTTACTGAACACTAGAGGCCATGAGCCAGAC 2219
Qy 2789 AATTAATGTATCTCCAGAGCTTGTGTAATTAAGCAGGCGACTGAGACACC 2848
Dh 2220 AATTAATGTATCTCCAGAGCTTGTGTAATTAAGCAGGCGACTGAGACACC 2279
Qy 2849 CGTCCGCTGTCTTCAACAGAAATCTTAAGAAAGAAATTAACAGAAAGAAAGCAAGA 2908
Dh 2280 CGTCCGCTGTCTTCAACAGAAATCTTAAGAAAGAAATTAACAGAAAGAAAGCAAGA 2339
Qy 2909 GGTTCCTGTGACACGCGCAACTTCAATGTCTCAGTGGAGCTCATTTCAAGGGCAAGACA 2968
Dh 2340 GGTTCCTGTGACACGCGCAACTTCAATGTCTCAGTGGAGCTCATTTCAAGGGCAAGACA 2399
Qy 2969 TTGAAAATGATGAATTCCAATCTGATACAGTATGACAGCTCATGTGCTCAACTTA 3028
Dh 2400 TTGAAAATGATGAATTCCAATCTGATACAGTATGACAGCTCATGTGCTCAACTTA 2459
Qy 3029 GGTGTGCGGTATGACAGCTGTATATGAGAGAGAGGCTTACCTTACATAGG 3088
Dh 2460 GGTGTGCGGTATGACAGCTGTATATGAGAGAGAGGCTTACCTTACATAGG 2519
Qy 3089 TTGCAGACAGCCCTGAGTTTCAAGAGTGTAAACAGAGGCTTGCCCTTTCAGCAACAGT 3148
Dh 2520 TTGCAGACAGCCCTGAGTTTCAAGAGTGTAAACAGAGGCTTGCCCTTTCAGCAACAGT 2579
Qy 3149 TCCAAATTCAGAGGCTTACAGGTCCTTACTCTCACAGGGGCTCCCAAGATGAAAC 3208
Dh 2580 TCCAAATTCAGAGGCTTACAGGTCCTTACTCTCACAGGGGCTCCCAAGATGAAAC 2639
Qy 3209 GACAAATGTGCTTTTATTTATTTATTTATTTGAGTGTCTGTATTTAAGAAATCAAA 3268
Dh 2640 GACAAATGTGCTTTTATTTATTTATTTATTTGAGTGTCTGTATTTAAGAAATCAAA 2699
Qy 3269 GTATTAACACATGCTTCTTCACTGACTTATTAATTAATCAATCAATCAATGATGAT 3328
Dh 2700 GTATTAACACATGCTTCTTCACTGACTTATTAATTAATCAATCAATCAATGATGAT 2759
Qy 3329 GCTGTGGTGTGACTTCTACTGACCGGTAGATTAACGTGTGTCTCCCAAGTGGTGG 3388
Dh 2760 GCTGTGGTGTGACTTCTACTGACCGGTAGATTAACGTGTGTCTCCCAAGTGGTGG 2819

OY	3389	GAAATATTTCACATCTGTCCAACCAAAAAAGATGCGTGTTTAGAGCATTGACACA	3448
Dd	2820	GAAATATTTCACATCTGTCCAACCAAAAAAGATGCGTGTTTAGAGCATTGACACA	2879
OY	3449	TATCTGCTTGGATTAAGAGACTTCTGAATTCCTAGGTGCGTTCGTTATCCATTGTG	3508
Dd	2880	TATCTGCTTGGATTAAGAGACTTCTGAATTCCTAGGTGCGTTCGTTATCCATTGTG	2929
OY	3509	GAAATTCATCTTGAATCCACTTGTCTATAATGTCATAGCANTAAGAAAATTTCTCAAGT	3568
Dd	2940	GAAATTCATCTTGAATCCACTTGTCTATAATGTCATAGCANTAAGAAAATTTCTCAAGT	2999
OY	3569	TTCCATGTGGGGTTCCTCGTAGCGAGCAATCTTTGSAATTTAAAGAAATTTAGAGA	3628
Dd	3000	TTCCATGTGGGGTTCCTCGTAGCGAGCAATCTTTGSAATTTAAAGAAATTTAGAGA	3059
OY	3629	ATATATTCATCTCTATAAAATGTTTTAAATATATACCMAACAGTGGCCCCCTGCATATAGTT	3688
Dd	3060	ATATATTCATCTCTATAAAATGTTTTAAATATATACCMAACAGTGGCCCCCTGCATATAGTT	3119
OY	3689	TTTCTGTTCGCACATGCAACCCATTACTTGGTACCTTAAAAACACACANTACTTATAGTC	3748
Dd	3120	TTTCTGTTCGCACATGCAACCCATTACTTGGTACCTTAAAAACACACANTACTTATAGTC	3179
OY	3749	CTGGGGATCGAAGATTCOAAAAATGSGATGCCGGAATGAAAAATCAAGGTGCAGCAGAGCT	3808
Dd	3180	CTGGGGATCGAAGATTCOAAAAATGSGATGCCGGAATGAAAAATCAAGGTGCAGCAGAGCT	3229
OY	3809	GTEGCTCTTGTGAAGGCTCTAGGGAGAMGCGGTTCTTTCGCATTTCAGGCTTCTAGAGG	3868
Dd	3240	GTEGCTCTTGTGAAGGCTCTAGGGAGAMGCGGTTCTTTCGCATTTCAGGCTTCTAGAGG	3299
OY	3869	CTGGCTTCATCTCCAGGCTTCAGTGGCTGTGTCAGCTTTTCTCACATGGCATCACTGTGA	3928
Dd	3300	CTGGCTTCATCTCCAGGCTTCAGTGGCTGTGTCAGCTTTTCTCACATGGCATCACTGTGA	3359
OY	3929	CACGTGGCCCTCCACATCTCCCTCTTTACTTTACAAAGGCCACAGGAATCCAGATATAT	3988
Dd	3360	CACGTGGCCCTCCACATCTCCCTCTTTACTTTACAAAGGCCACAGGAATCCAGATATAT	3419
OY	3989	CTCTCATCTTAAAGATCTTTCATCATCTCTGGAAGAGCCTTTGTGCCATGCAAGACAGATA	4048
Dd	3420	CTCTCATCTTAAAGATCTTTCATCATCTCTGGAAGAGCCTTTGTGCCATGCAAGACAGATA	3479
OY	4049	GCCACAGTGGGGATTAGACACAGACATCTTTGGGGTGTCTTATTTGGCTTACACAC	4108
Dd	3480	GCCACAGTGGGGATTAGACACAGACATCTTTGGGGTGTCTTATTTGGCTTACACAC	3539
OY	4109	CTTTCCTGCACACTAGTCTCCACAGAGAGGCTCTCAAATATGATCTGGCGACAGGGATGTTT	4168
Dd	3540	CTTTCCTGCACACTAGTCTCCACAGAGAGGCTCTCAAATATGATCTGGCGACAGGGATGTTT	3599
OY	4169	TGTTTTAGCTTGGCGACTTAACACTT-AAAAAAAAACCCAGATACAGAAATCTGGCCATGC	4227
Dd	3600	TGTTTTAGCTTGGCGACTTAACACTT-AAAAAAAAACCCAGATACAGAAATCTGGCCATGC	3659
OY	4228	TGGGGCTCACATTTCTACCTAGCAACACTGGCTGGAGCTGGGACACAGCTCTGCCCTTA	4287
Dd	3660	TGGGGCTCACATTTCTACCTAGCAACACTGGCTGGAGCTGGGACACAGCTCTGCCCTTA	3719
OY	4288	GAAAGGGTGTCCACTTACCAAGSTACACACAGCCACACACTAGGCCCTTATCTTCCACA	4347
Dd	3720	GAAAGGGTGTCCACTTACCAAGSTACACACAGCCACACACTAGGCCCTTATCTTCCACA	3779
OY	4348	ATGAGGCTTAGTGTCTTGTCTTCTATCATATCAATGAGGCCCTGCAAGGTTGCATTTATGTATG	4407
Dd	3780	ATGAGGCTTAGTGTCTTGTCTTCTATCATATCAATGAGGCCCTGCAAGGTTGCATTTATGTATG	3839
OY	4408	AAAAAGAAAGACATGGGATTAATCTCTAAATCAGGTAGTAGACATGAAACCAATATGTGTGC	4467
Dd	3840	AAAAAGAAAGACATGGGATTAATCTCTAAATCAGGTAGTAGACATGAAACCAATATGTGTGC	3899
OY	4468	TCACATTTACCCCTTTTCTTTTTTCTTTTTCTTTTTTCTTTTTTAAATGTAGACA	4527

[illegible]

QY 564 GAAAAATATTGACTGATGTGACGCCATGTCCTTTGGGAGGTGACCTTCAGCCCTC 623
DB 362 GAAAAATATTGACTGATGTGACGCCATGTCCTTTGGGAGGTGACCTTCAGCCCTC 421
QY 624 GACATGCTGTGCTGCTGCTCAACAGAACTTCATCTGGGATGTCAAACTCTAAGAG 683
DB 422 GACATGCTGTGCTGCTGCTCAACAGAACTTCATCTGGGATGTCAAACTCTAAGAG 481
QY 684 CATGCGTTAGAGTGCAGTTTTCATCCCTGCGCTGAGGACAGTCCGCTGGGAGAG 743
DB 482 CATGCGTTAGAGTGCAGTTTTCATCCCTGCGCTGAGGACAGTCCGCTGGGAGAG 541
QY 744 CTGCCCCAGACGAGTCACTCACTCCATCAGCGGCCAATGATGATCCAGCTGTGAGAT 803
DB 542 CTGCCCCAGACGAGTCACTCACTCCATCAGCGGCCAATGATGATGATCCAGCTGTGAGAT 601
QY 804 CGGAACCTTGTGAGCAATGAGCACTGTGTCCCGATCAAAATGCAAAAGAGTGAAT 863
DB 602 TGGAACTTGTGAGCAATGAGCACTGTGTCCCGATCAAAATGCAAAAGAGTGAAT 661
QY 864 GGCCTTACACCTCCATGTTCCACCCAGAAATGTCCTCGGCTTCAGCATTTGCAACCG 923
DB 662 GGCCTTACACCTCCATGTTCCACCCAGAAATGTCCTCGGCTTCAGCATTTGCAACCG 721
QY 924 CTGATCTATAAAAGCTGTGATCATGAGTGTGTTGAGGGTGAAGGCTCAGAAC 983
DB 722 CTGATCTATAAAAGCTGTGATCATGAGTGTGTTGAGGGTGAAGGCTCAGAAC 781
QY 984 CTGATCTATAAAAGCTGTGATCATGAGTGTGTTGAGGGTGAAGGCTCAGAAC 1043
DB 782 CTGATCTATAAAAGCTGTGATCATGAGTGTGTTGAGGGTGAAGGCTCAGAAC 841
QY 1044 TGTGCTTCTGCAACCTGCGGGGACAGCTCTCTCTCACTCACTCACTCACTG 1103
DB 842 TGTGCTTCTGCAACCTGCGGGGACAGCTCTCTCTCACTCACTCACTCACTG 901
QY 1104 TGAAGAGAGAGAGAGGCGGTTGATATCTACATCCGCGGCTCCACCAACCCCGAGT 1163
DB 902 TGAAGAGAGAGAGAGGCGGTTGATATCTACATCCGCGGCTCCACCAACCCCGAGT 961
QY 1164 GTTCAAGCTGAGAGACAAAGAGCTGAGAACATGGGAGGAGCTCAACCTCTCTGCA 1223
DB 962 GTTCAAGCTGAGAGACAAAGAGCTGAGAACATGGGAGGAGCTCAACCTCTCTGCA 1021
QY 1224 AGGCTGTGACCAAGATGCCCCAAAGTCCAGGATCTCTCGGCTGCACTTCCAAAGTTTGT 1283
DB 1022 AGGCTGTGACCAAGATGCCCCAAAGTCCAGGATCTCTCGGCTGCACTTCCAAAGTTTGT 1081
QY 1284 CCAACATCCCAAAATGAAGCAATTAATCTACGTGTGATGATGAATGAGCGAGC 1343
DB 1082 CCAACATCCCAAAATGAAGCAATTAATCTACGTGTGATGATGAATGAGCGAGC 1141
QY 1344 CATGTCACTGACCATGAGCAGCGGCCGTCAAAGAGCGCAAGTTTCCCTGGCTG 1403
DB 1142 CATGTCACTGACCATGAGCAGCGGCCGTCAAAGAGCGCAAGTTTCCCTGGCTG 1201
QY 1404 TTTCTGTGTCTAGAACTCTGCACTGCACTAGCAACCTCACTGATCTGATCTCAAA 1463
DB 1202 TTTCTGTGTCTAGAACTCTGCACTGCACTAGCAACCTCACTGATCTGATCTCAAA 1261
QY 1464 ACACAAAATCTCTCTCTTTGTGATGATCTGACACGTCTGTGATGAATGTGAAAAAAC 1523
DB 1262 ACACAAAATCTCTCTCTTTGTGATGATCTGACACGTCTGTGATGAATGTGAAAAAAC 1321
QY 1524 CATTAAGTGCAGACACCGGCTGCTGCAAAAGAAATCTTACTCACTCCAGGTGCCAG 1583
DB 1322 CATTAAGTGCAGACACCGGCTGCTGCAAAAGAAATCTTACTCACTCCAGGTGCCAG 1381
QY 1584 TGACATCTCTGACCTGCTGTGAGCTGATGATCTCTCTGAGAGTGTGTTGCCAA 1643
DB 1382 TGACATCTCTGACCTGCTGTGAGCTGATGATCTCTCTGAGAGTGTGTTGCCAA 1441
QY 1644 GGACAGGCTCAGGCTGTGCTGTGTCAGGCCAGAAAGCTGACAGCATACACACGAGAA 1703

DB 1442 GGACAGGCTCAGGCTGTGTGTGTCAGGCCAGAAACCTCAGCAGCATACACGAGAA 1501
QY 1704 GCCCTGCACACCAAGCTTACGATCACTCGTGGCAGTGGCATACCCAGGACAGCTGTA 1763
DB 1502 GCCCTGCACACCAAGCTTACGATCACTCGTGGCAGTGGCATACCCAGGACAGCTGTA 1561
QY 1764 CTTCGAGCTCTTGTGCCGGGAGGCTCTATCAAGCAGATCAGGTGAAGCAAAATCTC 1823
DB 1562 CTTCGAGCTCTTGTGCCGGGAGGCTCTATCAAGCAGATCAGGTGAAGCAAAATCTC 1621
QY 1824 GTTACCTCTTGCACCTTTGCCCCAGCTTCCAAAGAGGCTCCAGGAGGCTGTGAC 1883
DB 1622 GTTACCTCTTGCACCTTTGCCCCAGCTTCCAAAGAGGCTCCAGGAGGCTGTGAC 1681
QY 1884 GGTGCTCTTATACCTTATTTCAAGAGGAGGCTTTCACGCTGACCTGACACAAA 1943
DB 1682 GGTGCTCTTATACCTTATTTCAAGAGGAGGCTTTCACGCTGACCTGACACAAA 1741
QY 1944 AAGCAAGCTTACCTGAGACCCCACTGGAGCCGGGCTGCACTCCCTCACTCTGT 2003
DB 1742 AAGCAAGCTTACCTGAGACCCCACTGGAGCCGGGCTGCACTCCCTCACTCTGT 1801
QY 2004 GTCTGGAACATCAGCGTCCCAAGACAGTGTGCTGCTGATCTTTAAAGAGCG 2063
DB 1802 GTCTGGAACATCAGCGTCCCAAGACAGTGTGCTGCTGATCTTTAAAGAGCG 1861
QY 2064 GAGCGGCTGTGCTGCCAGACAGGCGCGCATTCATGATATCCAGAGAGAGGAGCCG 2123
DB 1862 GAGCGGCTGTGCTGCCAGACAGGCGCGCATTCATGATATCCAGAGAGAGGAGCCG 1921
QY 2124 GAGCTGAGAGATCTTACGCTGAGACAGATGTGCTCCCAAGCAAGCTTCCACATCA 2183
DB 1922 GAGCTGAGAGATCTTACGCTGAGACAGATGTGCTCCCAAGCAAGCTTCCACATCA 1981
QY 2184 CAGCTTGTGGTCAACATCTCTAATCTGACAGCCCAAGAGGCGCAAGCAGTGTGCT 2243
DB 1982 CAGCTTGTGGTCAACATCTCTAATCTGACAGCCCAAGAGGCGCAAGCAGTGTGCT 2041
QY 2244 CTTCGCTGACACTTACCCCAAGAGGCTGTGATGATCTGATCTGATCTGATCTGAT 2303
DB 2042 CTTCGCTGACACTTACCCCAAGAGGCTGTGATGATCTGATCTGATCTGATCTGAT 2101
QY 2304 GGGAGGTGAGTCTTACTGCTGTGCTGCTGCGGCTCATCTTGTGTGAAAAAGAA 2363
DB 2102 GGGAGGTGAGTCTTACTGCTGTGCTGCTGCGGCTCATCTTGTGTGAAAAAGAA 2161
QY 2364 GAAAAAGAGACAAACAGGCGCCGCTGTGGTATCTACATGCAAT 2414
DB 2162 AAAAAAAMAAACA---AGGGGCCCGCTGTGGGTATCTACAAATGCAACAT 2209

RESULT 6
AAS21324
ID AAS21324 standard; cDNA; 1376 BP.
XX
XX AAS21324;
AC
XX
DT 24-OCT-2001 (first entry)
XX
DE Human cDNA sequence encoding for PRO5773 polypeptide.
XX
XX Human secretory and transmembrane; PRO; mammalian; cancer; lung;
KW breast; prostate; cervical; tumour necrosis factor- α ; TNF- α ;
KW cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;
KW adipocyte; A-peptide; factor VIIA; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN WO200140466-A2.
XX
PD 07-JUN-2001.
XX

PF 01-DEC-2000; 2000MO-US32678.
 XX
 PR 01-DEC-1999; 99MO-US28301.
 PR 01-DEC-1999; 99MO-US28634.
 PR 02-DEC-1999; 99MO-US28531.
 PR 02-DEC-1999; 99MO-US28564.
 PR 02-DEC-1999; 99MO-US28565.
 PR 02-DEC-1999; 99MO-US28565.
 PR 16-DEC-1999; 99MO-US30095.
 PR 20-DEC-1999; 99MO-US30911.
 PR 20-DEC-1999; 99MO-US30911.
 PR 30-DEC-1999; 99MO-US31243.
 PR 06-JAN-2000; 2000MO-US00227.
 PR 11-FEB-2000; 2000MO-US00376.
 PR 18-FEB-2000; 2000MO-US04341.
 PR 18-FEB-2000; 2000MO-US04342.
 PR 22-FEB-2000; 2000MO-US04914.
 PR 24-FEB-2000; 2000MO-US05004.
 PR 24-FEB-2000; 2000MO-US05004.
 PR 01-MAR-2000; 2000MO-US05601.
 PR 20-MAR-2000; 2000MO-US07377.
 PR 21-MAR-2000; 2000MO-US07532.
 PR 30-MAR-2000; 2000MO-US08439.
 PR 17-MAY-2000; 2000MO-US13705.
 PR 22-MAY-2000; 2000MO-US14042.
 PR 30-MAY-2000; 2000MO-US14941.
 PR 02-JUN-2000; 2000MO-US15264.
 PR 10-NOV-2000; 2000MO-US30873.
 XX
 PA (GETH) GENENTECH INC.
 PI Baker KP, Betesini M, DeForge L, Desnoyers L, Filvaroff E, Gao W;
 PI Geritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
 PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;
 XX
 DR WPI: 2001-408281/43.
 XX P-PSDB: ANU12252.
 PT Isolated, secretory and transmembrane PRO polypeptide used to detect
 PT other PRO polypeptides, link bioactive molecules to cells expressing
 PT PRO polypeptides, and detect the presence of mammalian tumours e.g.
 PT lung, breast, prostate, cervical
 XX
 PS Claim 3; Fig 161; 813pp: English.
 CC AAS21244-AAS21518 encode for novel human secretory and transmembrane
 CC PRO polypeptides. The PRO polypeptides are useful to detect other
 CC PRO polypeptides, to link bioactive molecules to cells expressing
 CC PRO polypeptides, to modulate biological activities of cells expressing
 CC PRO polypeptides, and to detect the presence of mammalian lung, colon,
 CC breast, prostate, rectal, cervical or liver tumours by comparing PRO
 CC polypeptide expression in a cell sample to that in a control sample.
 CC Some of the 275 sequences are also useful to stimulate the release of
 CC tumour necrosis factor-alpha (TNF-alpha) from human blood, the
 CC proliferation or differentiation of chondrocytes, the proliferation or
 CC gene expression in pericyte cells, the release of proteoglycans from
 CC cartilage, the proliferation of inner ear utricular supporting cells or
 CC of T-lymphocytes, the release of a cytokine from peripheral blood
 CC monocytes (PBMCs), or the proliferation of endothelial cells. Some of
 CC the PRO polypeptides may modulate glucose or free fatty acid uptake by
 CC skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide
 CC to factor VIIa. The PRO polypeptides can be used in assays to identify
 CC molecules involved in binding interactions. The polynucleotides encoding
 CC PRO polypeptides can be used to generate probes, antisense RNA/DNA,
 CC transgenic or knock out animals and can be used in gene therapy.
 XX
 SQ Sequence 1376 BP; 311 A; 384 C; 352 G; 329 T; 0 other;
 Query Match 18.1%; Score 1117.8; DB 22; Length 1376;
 Best Local Similarity 99.8%; Pred. No. 1.1e-239;
 Matches 1119; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 190 CCAGGCGAGGCGACAGCTGCGCCGGGCTTGGGCGCCGCGGCTCCACCCTG 249
 DB 1 CCAAGGCGAGGCGACAGCTGCGCCGGGCTTGGGCGCCGCGGCTCCACCCTG 60
 QY 250 TTTTCCCAACCGAGGCGGCGTCCCGAGTCAATGCGCGGCTGAACTGCGGGTCTCT 309
 DB 61 TTTTCCCAACCGAGGCGGCGTCCCGAGTCAATGCGCGGCTGAACTGCGGGTCTCT 120
 QY 310 ATCGACCTGTAGGGGTTTCTGCTGCTGCGGGTGGGCGGCTTGGCGGCGGAGAGCT 369
 DB 121 ATCGACCTGTAGGGGTTTCTGCTGCTGCGGGTGGGCGGCTTGGCGGCGGAGAGCT 180
 QY 370 TTTGAGATTGCTGTGCGACGAGAGAAACCATATACAGTTCTCTAAAGCTGGGAGCCCG 429
 DB 181 TTTGAGATTGCTGTGCGACGAGAGAAACCATATACAGTTCTCTAAAGCTGGGAGCCCG 240
 QY 430 ACTCTGCTGGCAAAACCTGTTACATGTCATTTCTAAAGACATATTAACATGTTGTCC 489
 DB 241 ACTCTGCTGGCAAAACCTGTTACATGTCATTTCTAAAGACATATTAACATGTTGTCC 300
 QY 490 ATCAAGTCTGGAGAAAGATAGTCTTACCTTACCTTACCTGCGACAGTCTGAGATCACTT 549
 DB 301 ATCAAGTCTGGAGAAAGATAGTCTTACCTTACCTTACCTGCGACAGTCTGAGATCACTT 360
 QY 550 GTCATGAGATCCAGAAATATATGACTGATGTCAGGCGCCATGTCCTTGGGAGGTT 609
 DB 361 GTCATGAGATCCAGAAATATATGACTGATGTCAGGCGCCATGTCCTTGGGAGGTT 420
 QY 610 CAGCTTCAGCCCTCGACATGTTGTTGCTACCTTACCTTACCTTACCTTACCTTACCTT 669
 DB 421 CAGCTTCAGCCCTCGACATGTTGTTGCTACCTTACCTTACCTTACCTTACCTTACCTT 480
 QY 670 AAAGCTCATTAAGATCACTGCTTTAGCTGAGCTGAGCTTTTCCATCCCTGCTGAGGCGATC 729
 DB 481 AAAGCTCATTAAGATCACTGCTTTAGCTGAGCTGAGCTTTTCCATCCCTGCTGAGGCGATC 540
 QY 730 GGTCCGGGTGAGAGTGGCCAGAGGAGTCACTCACTCATCGAGCGGCGGAATCGATGCC 789
 DB 541 GGTCCGGGTGAGAGTGGCCAGAGGAGTCACTCACTCATCGAGCGGCGGAATCGATGCC 600
 QY 790 ACCGTGTGAGATCGGAACCTTCTGACGAAATGAGCACTGTGTCCCGGATCAAGATGCA 849
 DB 601 ACCGTGTGAGATCGGAACCTTCTGACGAAATGAGCACTGTGTCCCGGATCAAGATGCA 660
 QY 850 GAAGGATGAATATGCGCTTACACCTCCATGCTTCCACCCAGAAATGTCGCGCTTC 909
 DB 661 GAAGGATGAATATGCGCTTACACCTCCATGCTTCCACCCAGAAATGTCGCGCTTC 720
 QY 910 AGCATGCAAAACCGCTCATCTATAAAGCTGTGTGATCATGAGTGTGAGGGT 969
 DB 721 AGCATGCAAAACCGCTCATCTATAAAGCTGTGTGATCATGAGTGTGAGGGT 780
 QY 970 GAAGGCTCAACACCTGATGTGTGCACTACCAAGAGCTTCCCTGAGATGAGCTC 1029
 DB 781 GAAGGCTCAACACCTGATGTGTGCACTACCAAGAGCTTCCCTGAGATGAGCTC 840
 QY 1030 ATGACGTGACAGTTTGTCTCTGTGCAACCTGCGGCGAGCTGTCTCTCAACTTC 1089
 DB 841 ATGACGTGACAGTTTGTCTCTGTGCAACCTGCGGCGAGCTGTCTCTCAACTTC 900
 QY 1090 AACCTTCCAACTGTAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1149
 DB 901 AACCTTCCAACTGTAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 960
 QY 1150 ACCAACCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1209
 DB 961 ACCAACCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1020
 QY 1210 AACCTTCTGCAAGGCTGTGACCAAGATGCCAAAGTCCAGAGGATCTCTCGGCTGCA 1269
 DB 1021 AACCTTCTGCAAGGCTGTGACCAAGATGCCAAAGTCCAGAGGATCTCTCGGCTGCA 1080
 QY 1270 TTCCAAGTTTGTGTCACATCCAAATGAAGCAATRA 1310

DB 1081 TTCCAGTTTGTGTCACACATCCACAAATGAAAGCAGTGA 1121
|||||
RESULT 7
ABV30293
ID ABV30293 standard; cDNA; 1031 BP.
XX
XX ABV30293;
AC
XX 16-SEP-2002 (first entry)
DT
XX
XX Human prostate expression marker cDNA 30284.
DE
XX
XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KM pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
PN WO200160860-A2.
XX
XX 23-AUG-2001.
PD
XX
XX 20-FEB-2001; 2001MO-US05171.
PF
XX
XX 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-235281P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
PI
XX Schlegel R, Endege WO, Monahan JE;
XX
XX WPI; 2001-662795/76.
DR
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
XX Claim 1; Page 6580; 11750pp; English.
PS
XX
XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
CC
XX
XX Sequence 1031 BP; 272 A; 255 C; 212 G; 263 T; 29 other;
SQ
Query Match 11.7%; Score 718; DB 23; Length 1031;
Best Local Similarity 96.5%; Pred. No. 2.1e-150;
Matches 745; Conservative 0; Mismatches 25; Indels 2; Gaps 1;

QY 4619 TCACCTCCCAAAATAGCTGGATCATCTGGCAAAACCACCATGCCAGCTAATTTGTAT 4678
|||||
DB 150 TCAGCTCCCAAAATAGCTGGATCATCTGGCAAAACCACCATGCCAGCTAATTTGTAT 209
QY 4679 TTTTGTAGAGACAGGTTTACCATTTGGCCAGGCTGCTCAGCTCCCTGGGCTCAA 4738
|||||
DB 210 TTTTGTAGAGACAGGTTTACCATTTGGCCAGGCTGCTCAGCTCCCTGGGCTCAA 269
QY 4739 GCATTCCTCGCTGGCTGGCTCCCAAGTGTGGATTTACAGATGTAGAGCAGCCGATCC 4798
|||||
DB 270 GCATTCCTCGCTGGCTGGCTCCCAAGTGTGGATTTACAGATGTAGAGCAGCCGATCC 329
QY 4799 AGCCACACACCTCATTTATACCAATTAATCTGCCATTAATCTGTGAGCTTTGGCTTCTC 4858
|||||
DB 330 AGCCACACACCTCATTTATACCAATTAATCTGCCATTAATCTGTGAGCTTTGGCTTCTC 389
QY 4859 ACCCTGCTCTGATCTGGAAGAGAGAGGATTTAGTTATAGCTTGTACAGACAGTCCCAAG 4918
|||||
DB 390 ACCCTGCTCTGATCTGGAAGAGAGGATTTAGTTATAGCTTGTACAGACAGTCCCAAG 449
QY 4919 TTCAATATTTCTGGGGCAAAACTTCCTCAAAAATTAATATGACTTCATTTGATTCAT 4978
|||||
DB 450 TTCAATATTTCTGGGGCAAAACTTCCTCAAAAATTAATATGACTTCATTTGATTCAT 509
QY 4979 GAATTCACCTTGGAAATGACACCGCTCAACTTGTTCACATGGCATTAATGAAGAAATTT 5038
|||||
DB 510 GAATTCACCTTGGAAATGACACCGCTCAACTTGTTCACATGGCATTAATGAAGAAATTT 569
QY 5039 TATAGTCTCTTAATGAGCGGTGTACTGCAAGACCTTGAACACTTTCCAGAGATAGAT 5098
|||||
DB 570 TATAGTCTCTTAATGAGCGGTGTACTGCAAGACCTTGAACACTTTCCAGAGATAGAT 629
QY 5099 ATTTAAGTCATGACCTTGGGCTTGTGACCTATGACACTTCCCTTGTGAAGTCTGCTCCT 5158
|||||
DB 630 ATTTAAGTCATGACCTTGGGCTTGTGACCTATGACACTTCCCTTGTGAAGTCTGCTCCT 689
QY 5159 GCCCAGTAGACCTTGGCTTGTGACCGAGATGCTGACCTGCAATGAAGGCCAAAGAG 5218
|||||
DB 690 GCCCAGTAGACCTTGGCTTGTGACCGAGATGCTGACCTGCTT--AGGCCAAAGAGAG 747
QY 5219 GCGTGGCGCTTCCCTCCCTCAGTGAAGAGCCCTTATTTGATTCAGTGTG 5270
|||||
DB 748 GCTGCAAGTTTCCCTCCCTCAGTGAAGAGCCCTTATTTGATTCAGTGTGAG 799
RESULT 8
AAH08379
ID AAH08379 standard; cDNA; 659 BP.
XX
XX AAH08379;
AC
XX
XX 26-JUN-2001 (first entry)
DT
XX
XX Human cDNA clone (5'-primer) SEQ ID NO:5214.
DE
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
KM
XX
XX Homo sapiens.
OS
XX
XX EP1074617-A2.
PN
XX
XX 07-FEB-2001.
PD
XX
XX 28-JUL-2000; 2000EP-0116126.
PF
XX
XX 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI: 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
XX Claim 1: SEQ ID 5214; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH0166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
XX Sequence 659 BP; 150 A; 187 C; 168 G; 151 T; 3 other:
SQ
Query Match 10.2%; Score 630.8; DB 22; Length 659;
Best Local Similarity 98.9%; Pred. No. 5,4e-131;
Matches 654; Conservative 0; Mismatches 5; Indels 2; Gaps 2;

OY 1049 TTCCTGACACCTGGCGGCGAGCTGCTTCTCACTTCAACCTCTCCAACTGTGAGA 1108
DB 480 TTCTCTCACACCTGCGGGCGACGCTCTCTCTCACTTCAACCTCTCCAACTGTGAGA 539
OY 1109 GGAAGAGGAGCGGGTTGAATACATCCCGGGCTCCACACCAACCCGAGGTGTCA 1168
DB 540 GGAAGAGGAGCGGGTTGAATACATCCCGGGCTCCACACCAACCCGAGGTGTCA 599
OY 1169 AGCTGGAGGACAAAGCAGCTGGGACATGGCGGGGAACTTCAACTCTCTGTGCAAGGCT 1228
DB 600 A-CTGGAGGACAAAGCAGCTGGGACATGGCGGGGAACTTCAACTCTGTGCAAGGCT 658
OY 1229 G 1229
DB 659 G 659
RESULT 9
ID AAA78075 standard; cDNA; 544 BP.
XX AAA78075;
AC
XX
XX
XX 14-NOV-2000 (first entry)
DT
XX
XX cDNA encoding human colon tumour polypeptide. SEQ ID NO:362.
DE
XX
XX Human colon tumour polypeptide; tumour antigen; cancer; vaccine;
KW Immunotherapy; diagnosis; progression; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200037643-A2.
PN
XX
XX 29-JUN-2000.
PD
XX
XX 23-DEC-1999; 99WO-US30909.
PF
XX
XX 23-DEC-1998; 98US-0221298.
PR
XX 02-JUL-1999; 99US-0347496.
PR 22-SEP-1999; 99US-0401064.
PR 19-NOV-1999; 99US-0444242.
PR 02-DEC-1999; 99US-0454150.
XX
XX (CORI-) CORIXA CORP.
PA
XX
XX Xu J, Lodes MJ, Secretist H, Benson DR, Meagher MJ, Stolk J;
PI Wang T, Yqdu J;
PI
XX
XX WPI: 2000-442671/38.
DR
XX
XX New colon tumor polypeptides used to inhibit the development of cancer,
PT especially colon cancer, and for diagnosing and monitoring the
PT progression of the cancer -
PT
XX
XX Claim 1: Page 196; 229pp; English.
PS
XX
XX Sequences AAA77722-A78199 represent 478 cDNAs encoding proteins or
CC portions of proteins which are associated with human colon tumours.
CC The invention also specifically discloses 8 human colon tumour proteins
CC (AAH1897-B11904). The nucleic acids, the polypeptides they encode, and
CC antigen presenting cells (APCs), preferably dendritic cells) expressing
CC such polypeptides may be used in vaccines that target tumour cells,
CC especially colon tumour cells, thereby inhibiting the development of
CC cancer. T-cells specific for the polypeptide expressed by the APC are
CC used to remove tumour cells from biological samples, especially blood or
CC fractions thereof. The sample or the isolated T-cells specific for the
CC polypeptide can then be used to inhibit cancer development. CD4+ and/or
CC CD8+ T-cells from a patient may be incubated with a polypeptide or
CC nucleic acid of the invention, or an APC expressing such a polypeptide,
CC to cause the proliferation of specific T-cells. The T-cells can be
CC cloned and then administered back to the patient to inhibit cancer

CC development. Nucleic acids encoding the polypeptides and antibodies
CC against the polypeptides may be used to determine the expression level
CC of a tumour protein of the invention, and therefore to determine whether
CC cancer cells are present. Such diagnostic methods may also be used to
CC monitor the progression of a cancer by repeating the processes at time
CC intervals, and comparing the current result to previous results. The
CC present sequence represents a cDNA encoding a human colon tumour
CC polypeptide.
CC
XX Sequence 544 BP; 138 A; 123 C; 113 G; 170 T; 0 other;
SQ
Query Match 8.8%; Score 542.4; DB 21; Length 544;
Best Local Similarity 99.8%; Pred. No. 2.8e-111;
Matches 543; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3068 CCTGAGTCACCATGACATAGAGGTTGACAGAGCCCTGGATTTCAGAGTTTAAACAGAGCT 3127
DB 1 CCTGAGTCACCATGACATAGAGGTTGACAGAGCCCTGGATTTCAGAGTTTAAACAGAGCT 60
QY 3128 TGGCCCTTTCAGAGCAACAGCTTCCAAATTCAGAGAGCCCTGAGCTTCCAC 3187
DB 61 TGGCCCTTTCAGAGCAACAGCTTCCAAATTCAGAGAGCCCTGAGCTTCCAC 120
QY 3188 TGGGAGTCCAGAGTGAAGAGCAATGCGCTTTTATATTATTATTGCGTCT 3247
DB -121 TGGGAGTCCAGAGTGAAGAGCAATGCGCTTTTATATTATTATTGCGTCT 180
QY 3248 GTGTTATTAAAGAGATCAAAATGTTAAACACCTAGCTCTTTTCCAGCTTACTATATA 3307
DB 181 GTGTTATTAAAGAGATCAAAATGTTAAACACCTAGCTCTTTTCCAGCTTACTATATA 240
QY 3308 CTCTACTAACTGTTGGATGCGTGGTGTGACTTCTACTGACCCCTGATTAACGTG 3367
DB 241 CTCTACTAACTGTTGGATGCGTGGTGTGACTTCTACTGACCCCTGATTAACGTG 300
QY 3368 TGCCGTGCCCCAGAGTGGTGGGAATTAATTAACAATCTGTCCAAACAGAGAAATGCTG 3427
DB 301 TGCCGTGCCCCAGAGTGGTGGGAATTAATTAACAATCTGTCCAAACAGAGAAATGCTG 360
QY 3428 TGTGTGAGCAGCATGACATATATCTGTTGATTAAGAGACTTCCGATTTCTAGCTG 3487
DB 361 TGTGTGAGCAGCATGACATATATCTGTTGATTAAGAGACTTCCGATTTCTAGCTG 420
QY 3488 GTTGTGAGTATCCATTTGGAATTCATCTGTAATCCCATTTCTATATAGTCTAGCA 3547
DB 421 GTTGTGAGTATCCATTTGGAATTCATCTGTAATCCCATTTCTATATAGTCTAGCA 480
QY 3548 ATAGAGAAATTCCTCAAGTTTCCATGCGGTTCTCTAGTCGAGCAATACTTGAC 3607
DB 481 ATAGAGAAATTCCTCAAGTTTCCATGCGGTTCTCTAGTCGAGCAATACTTGAC 540
QY 3608 ATTT 3611
DB 541 ATTT 544
RESULT 10
AA128813
ID AA128813 standard; cDNA; 544 BP.
XX
AC AA128813;
XX
DT 12-OCT-2001 (first entry)
XX
DE Colon tumour related determined cDNA sequence for clone 25928.
XX
KW Human; immunotherapy; diagnosis; colon cancer; colon tumour;
KW Immunogenic; gene therapy; vaccine; colonic cancer; ss.
XX
OS Homo sapiens.
XX
PN W0200149716-A2.

PD 12-JUL-2001.
XX
PF 29-DEC-2000; 2000NO-US35596.
XX
PR 30-DEC-1999; 99US-0476296.
XX
PR 10-JAN-2000; 2000US-0480321.
PR 15-FEB-2000; 2000US-0504629.
PR 06-MAR-2000; 2000US-0519444.
PR 19-MAY-2000; 2000US-0575251.
PR 29-JUN-2000; 2000US-0609448.
PR 28-AUG-2000; 2000US-0649811.
XX
PA (CORI-) CORIXA CORP.
PI Xu J, Lodes M, Secret H, Benson DR, Meagher M, Stolk JA;
PI King GE, Wang T, Jiang Y;
XX
DR WPI; 2001-441847/47.
XX
PT Colon tumor associated proteins and nucleic acids useful for the
PT prevention, diagnosis and treatment of colonic cancer -
XX
PS Claim 2; Page 233; 472pp; English.
XX
CC The present invention describes colon tumour associated proteins (I) and
CC the polynucleotides (II) that encode them. (I) have cytostatic activity.
CC (I) and (II) can be used in gene therapy and vaccine production. (I) and
CC (II) may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate colon tumour associated protein (TCAP)
CC expression, such as colonic cancer. For example, (I) and (II) may be
CC used to treat disorders associated with decreased expression by
CC rectifying mutations or deletions in a patient's genome that affect the
CC activity of TCAPs by expressing inactive proteins or to supplement the
CC patients own production of them. Additionally, (II) may be used to
CC produce the TCAP proteins, by inserting the nucleic acids into a host
CC cell culturing the cell to express the protein. (II) and its
CC complementary sequences may also be used as DNA probes in diagnostic
CC polymerase chain reaction (PCR) and hybridisation assays to detect and
CC quantitate the presence of similar nucleic acids in samples, and
CC therefore which patients may be in need of restorative therapy. (I) may
CC also be used as antigens in the production of antibodies against TCAPs
CC and in assays to identify modulators of TCAP expression and activity.
CC Anti-(I) antibodies and antagonists may also be used to down regulate
CC TCAP expression and activity. The anti-(I) antibodies may also be used
CC as diagnostic agents for detecting the presence of TCAPs in samples
CC (e.g. by enzyme linked immunosorbant assay (ELISA)). AA128460 to AA129512
CC and AA124494 to AA124523 represent nucleotide and amino acid sequences
CC given in the exemplification of the present invention.
XX
SQ Sequence 544 BP; 138 A; 123 C; 113 G; 170 T; 0 other;
Query Match 8.8%; Score 542.4; DB 22; Length 544;
Best Local Similarity 99.8%; Pred. No. 2.8e-111;
Matches 543; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3068 CCTGAGTCACCATGACATAGAGGTTGACAGAGCCCTGGATTTCAGAGTTTAAACAGAGCT 3127
DB 1 CCTGAGTCACCATGACATAGAGGTTGACAGAGCCCTGGATTTCAGAGTTTAAACAGAGCT 60
QY 3128 TGGCCCTTTCAGAGCAACAGCTTCCAAATTCAGAGAGCCCTGAGCTTCCAC 3187
DB 61 TGGCCCTTTCAGAGCAACAGCTTCCAAATTCAGAGAGCCCTGAGCTTCCAC 120
QY 3188 TGGGAGTCCAGAGTGAAGAGCAATGCGCTTTTATATTATTATTGCGTCT 3247
DB 121 TGGGAGTCCAGAGTGAAGAGCAATGCGCTTTTATATTATTATTGCGTCT 180
QY 3248 GTGTTATTAAAGAGATCAAAATGTTAAACACCTAGCTCTTTTCCAGCTTACTATATA 3307
DB 181 GTGTTATTAAAGAGATCAAAATGTTAAACACCTAGCTCTTTTCCAGCTTACTATATA 240
QY 3308 CTCTACTAACTGTTGGATGCGTGGTGTGACTTCTACTGACCCCTGATTAACGTG 3367

QY	5701	AATCTGTGGCATTTGGCAGCAAGGAAAGGTCCTCAAGGAAAGTCAACTGGGACAAAGCA	5760
Db	443	AATCTGTGGCATTTGGCAGCAAGGAAAGTCTCAAGGAAAGTCAACTGGGACAAAGCA	384
QY	5761	CAAGCCACCGACATGGCCCTGGTAAAGGTTAGCAGACTGTGTGTGGATCTGCAGT	5820
Db	383	CAAGCCACCGACATGGCCCTGGTAAAGGTTAGCAGACTGTGTGTGGATCTGCAGT	324
QY	5821	GCTTCACTGGCAATTAATTATTCATTGACAGTACTTTTAGTGGCATTTTATTCATTTTC	5880
Db	323	GCTTCACTGGCAATTAATTATTCATTGACAGTACTTTTAGTGGCATTTTATTCATTTTC	264
QY	5881	CTGTGCTTTAAATTAACAAATGTACCAAAAAACAAGATCAAGCTGTTTAAAGCTTCGG	5940
Db	263	CTGTGCTTTAAATTAACAAATGTACCAAAAAACAAGATCAAGCTGTTTAAAGCTTCGG	204
QY	5941	CTACTTGTCCCTGCTGTTCACTAGTAGAGGCCCGGTTTCCAGTTTGTGACTGTGACAGCTC	6000
Db	203	CTACTTGTCCCTGCTGTTCACTAGTAGAGGCCCGGTTTCCAGTTTGTGACTGTGACAGCTC	144
QY	6001	AGCATGGGCTTCAGCAGATGCTGTCTTAATTGTGTGATGATACAGAAACCCAGGCTTTGGG	6060
Db	143	AGCATGGGCTTCAGCAGATGCTGTCTTAATTGTGTGATGATACAGAAACCCAGGCTTTGGG	84
QY	6061	ATCAAGAGTCTTTCCTCTCTCATTTGGATGGCCGTGCACGTGTGAAAGCAGATGTTTGTCTC	6120
Db	83	ATCAAGAGTCTTTCCTCTCTCATTTGGATGGCCGTGCACGTGTGAAAGCAGATGTTTGTCTC	24
QY	6121	GGAATATTAATAATATAGTCTTGG 6143	
Db	23	GGAATATTAATAATATAGTCTTGG 1	
RESULT 14			
ID	AAZ80286/C		
ID	AAZ80286 standard; cDNA; 636 BP.		
XX	AAZ80286;		
XX	07-APR-2000 (first entry)		
DE	Human colon cancer cell line SW480 cDNA clone SEQ ID NO:370.		
KW	Human; gene expression product; diagnosis; tumour; colon cancer;		
KW	colorectal adenocarcinoma; cell line SW480; cell proliferation;		
KW	cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;		
XX	hyperplasia; ds.		
OS	Homo sapiens.		
PN	WC9964576-A2.		
XX	16-DEC-1999.		
XX	09-JUN-1999; 99WO-IB01062.		
PR	10-JUN-1998; 98US-0088801.		
PA	(FARB) BAYER CORP.		
PI	Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;		
PI	Carroll E, Catino TJ, Dertl A, Ford DM, Lewis ME, Monahan JE;		
PI	Schlegel R;		
DR	WPI; 2000-087220/07.		
PT	Novel nucleic acids, used to develop products for the diagnosis and		
PT	treatment of disorders involving unwanted cell proliferation,		
PT	particularly cancers, especially colon cancer		
XX	Claim 15; Page 280; 469pp: English.		
PS	AAZ79917 to AAZ80766 represent double stranded cDNA clones isolated from		
CC			

Query Match	7.9%	Score 489.2	DB 21	Length 636
Best Local Similarity	90.3%	Fred. No. 2.3e-99		
Matches 577	Conservative 0	Mismatches 54	Indels 8	Gaps 6
5270	GGAGCCCTAGAGCCCTCCATCTTCGACATATTCACACCTCCACCCCTTCGACAGAGACT	5329		
Db	636	GGAAGCCANNANNCCTTCATTTTAAANAATTCACAACCTT--CAAGCCCTTCCAAAGAGAT	579	
5330	AGGTGCCCCCTGCATTCACCCCAAGGATGGATGGCCCTTCCTT--AGGCTGGCTACCTTG--TC	5386		
Db	578	A-GNCCCTTCNGANTCCNCCCAAGAGGATGGCTTCCTTAAAGNNGGTNACTGTGATNC	520	
5387	ACCATCACCAGCATTCACATGTTTCCCTGCAAGAGACACACGTCGGCATTTTCCCTTC--AATG	5445		
Db	519	ACNATCAGGGAATNNCTGGTGTCTTGCCAGNAGNCCCAAGGGCCATTTTCCCTTCAANA	460	
5446	AGGGCTCAAAATCTCCGAGCAAGTTGCGTGCCTC--TGAGCAGCATTTTCTGAGAGCTCT	5504		
Db	459	AGGGCTCAAAATCTCCGAGCAAGTTGCGTGCCTTAAAGNCCAGATTTTCCCTGAAAGCTG	400	
5505	GCCTCAGTGAAGGGGCCAGCCCTGAGGAACCCCTGGCTCTTTTCTTTAAAGCCAGGCC	5564		
Db	399	CCTCAAGTAAAGGGGCCAGCTTGAGGAACCTGGCTCTTTTCTTTAAAGCCAGGCC	340	
5555	ACTTACATATAAATCATTTTCAGGGTCACTGGAACAGTGAAGTGCATTTTGTGAAGCTTC	5624		
Db	339	ACTTACATATAAATCATTTTCAGGGTCACTGGAACAGTGAAGTGCATTTTGTGAAGCTTC	280	
5625	TGCATGCCAGCCCATCTGCTCATCCAGTGGTGTGCGATCCCTACGAGAGAGCCAGCGCA	5684		
Db	279	TGCATGCCAGCCCATCTGCTCATCCAGTGGTGTGCGATCCCTACGAGAGAGCCAGCGCA	220	
5685	TGCAGAGCTGTGCTCTTAATGCTGTGGTCAATTCACAGAAAGGAAAGGTCTCAAGGAAG	5744		
Db	219	TGCAGAGCTGTGCTCTTAATGCTGTGGTCAATTCACAGAAAGGAAAGGTCTCAAGGAAG	160	
5745	TCACTGGGACAGACAAAGCCACCGGACATGCGCTTGTAAAGTTAGCAGACTGGTG	5804		
Db	159	TCACTGGGACAGACAAAGCCACCGGACATGCGCTTGTAAAGTTAGCAGACTGGTG	100	
5805	TGTGTGAAATTCGACATGCTTCACGTGGAATATTTATTCATGCAATCTTTTAGGTG	5864		
Db	99	TGTGTGAAATTCGACATGCTTCACGTGGAATATTTATTCATGCAATCTTTTAGGTG	40	
5865	GCATTTATTCATTTCTGCTGTCTTAAATTAACAATGT	5903		
Db	39	GCATTTATTCATTTCTGCTGTCTTAAATTAACAATGT	1	
RESULT 15				
AAS64483				
ID	AAS64483	standard; cDNA; 487 BP.		
XX	AAS64483;			
XX				
XX	13-FEB-2002 (first entry)			
XX				
XX	DNA encoding novel human diagnostic protein #287.			
XX				
XX	Human; chromosome mapping; gene mapping; gene therapy; forensic;			
XX	food supplement; medical imaging; diagnostic; genetic disorder; ss.			

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 11, 2002, 09:32:51 : Search time 63 Seconds
(without alignments)
1768.215 Million cell updates/sec

Title: US-09-899-569a-4

Perfect score: 4394
Sequence: 1 MAGLNCGVSTALLGVLLGA.....SSKDTDPILNTQPMERAE 836

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4394	100.0	836	23	AA049641
2	3634.5	82.7	709	21	AA049645
3	3467	78.9	749	23	AA049640
4	3411	77.6	649	22	AA049558
5	2155	49.0	443	21	AA049592
6	1779	40.5	343	22	AA049252
7	827	18.8	159	21	AA049153
8	775	17.6	143	22	AB000296
9	616	14.0	121	22	AA023969
10	523	11.9	167	22	AA049483

11	390	8.9	116	22	AA090342	Human immune/haema
12	140.5	3.2	3623	20	AA072020	Rat cubilin proteol
13	126	2.9	595	22	AA070070	Human secreted pro
14	126	2.9	595	23	AA065509	Human albumin fusi
15	119	2.7	3095	23	AA020788	Rat C3b/C4b comple
16	118.5	2.7	1406	21	AA042916	Human ORFX ORF2680
17	118	2.7	1733	23	AA060623	Human cubilin prot
18	118	2.7	2118	22	AA023416	Novel human diagno
19	118	2.7	3631	22	AA028277	Novel human diagno
20	117.5	2.7	774	22	AA064637	Drosophila melano
21	117.5	2.7	774	22	AA038959	Drosophila G-prote
22	117.5	2.7	1464	22	AA071111	Drosophila melano
23	117.5	2.7	1719	17	AA092100	Human RIZ allele D
24	117	2.7	1045	23	AA091371	Herbicideally activ
25	117	2.7	2476	20	AA067738	P19 p105 zona pell
26	116.5	2.7	4215	22	AA059065	Drosophila melano
27	116	2.6	1890	23	AA060622	Human cubilin prot
28	114	2.6	477	23	AA023674	SIG protein. Unid
29	113.5	2.6	838	22	AA070548	Drosophila melano
30	113	2.6	3329	21	AA056029	Human BRCA2 tumou
31	112.5	2.6	433	17	AA094454	Oligodendrocyte-my
32	112.5	2.6	433	21	AA057091	Human oligodendroc
33	112.5	2.6	581	22	AA012455	Human bone marrow
34	112.5	2.6	1770	22	AA039102	Human polypeptide
35	112.5	2.6	1805	22	AA040410	Novel human diagno
36	112	2.5	736	23	AA057361	Mouse ischaemic co
37	112	2.5	1848	22	AA040888	Human polypeptide
38	111.5	2.5	728	22	AA085060	Human serine prote
39	111.5	2.5	1393	22	AA078991	Human protein SEQ
40	110.5	2.5	771	23	AA068311	Human semaphorin G
41	110.5	2.5	870	23	AA068303	Human semaphorin G
42	110.5	2.5	942	23	AA091468	Herbicideally activ
43	110.5	2.5	1694	21	AA044661	Rat laminin 5 poly
44	110.5	2.5	1719	21	AA012029	Human Rb-interacti
45	110.5	2.5	1719	21	AA012113	Human Rb-interacti

ALIGNMENTS

RESULT 1
AA049641
ID AA049641 standard: Protein: 836 AA.
AA049641:
17-MAY-2002 (first entry)
Human tumour-associated antigen B345 protein SPQ ID NO 4.
Tumour-associated antigen; human; B345; cytosolic; cell communication;
cell interaction; signal transduction; metastasis; cancer; colon;
immunotherapy; carcinoma; lung; diagnosis.
Homo sapiens.
WO200204508-A1.
17-JAN-2002.
05-JUL-2001; 2001WO-EP07705.
07-JUL-2000; 2000DE-1033080.
PR 19-APR-2001; 2001DE-1019294.
(BOEH) BOEHRINGER INGELHEIM INT GMBH.
Schweifer N, Scherl-Mostageer M, Sommergruber W, Abseher R;
WPI; 2002-171704/22.
N-PSDB; ABA99507.
New tumour-associated antigen B345, useful for diagnosis and

PT immunotherapy of tumors, also related nucleic acid and antibodies -
 XX
 XX Claim 1; Page 85-88; 102pp; German.
 CC This invention describes a novel tumour-associated antigen, designated
 CC B345 which has cytostatic activity. B345 is involved in communication,
 CC interaction and/or signal transduction with extracellular components and
 CC ligands, especially in the metastatic potential of cancers, particularly
 CC of the colon. B345 or its immunogenic fragments, also the DNA that
 CC encodes it, are useful for immunotherapy of cancer, particularly
 CC carcinoma of lung or colon. Antibodies raised against B345 are useful for
 CC treatment and diagnosis of cancers that are associated with B345
 CC expression, including their use for targeted delivery of cytotoxic or
 CC radioactive agents. Probes derived from B345 can be used to detect
 CC tumour-specific mutations in the B345 sequence, and can be used to screen
 CC for B345 specific modulators. This sequence represents a human B345
 CC tumour-associated antigen described in the invention.
 CC
 XX
 SQ Sequence 836 AA;
 Query Match 100.0%; Score 4394; DB 23; Length 836;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 836; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MAGLNCVSTALLGVLILGAAALPGCAEALPRESNTVLKIGTPPLAKPCYIV 60
 DB 1 MAGLNCVSTALLGVLILGAAALPGCAEALPRESNTVLKIGTPPLAKPCYIV 60
 QY 61 SKRHITMLSTKSGRIYFTFSCSPENHFYIEIOKNIDKSGCPCEGVOLOPSTSLPT 120
 DB 61 SKRHITMLSTKSGRIYFTFSCSPENHFYIEIOKNIDKSGCPCEGVOLOPSTSLPT 120
 QY 121 LNRTFMDVAKHSIGLELOFSIPRLKIOIGPBGSCPDGVTSHISGRIDATVVRIGTFCSN 180
 DB 121 LNRTFMDVAKHSIGLELOFSIPRLKIOIGPBGSCPDGVTSHISGRIDATVVRIGTFCSN 180
 QY 181 GTVSRKIMQGVKMAHLPMFHPRANVSGESIANRSTIKRLCTIESVREGSGSATLMSANT 240
 DB 181 GTVSRKIMQGVKMAHLPMFHPRANVSGESIANRSTIKRLCTIESVREGSGSATLMSANT 240
 QY 241 PEGPPEDELTMQFVPAHLRASVFLNFMLSNCERKEEVEYIIPGSTNPEYFKEDK 300
 DB 241 PEGPPEDELTMQFVPAHLRASVFLNFMLSNCERKEEVEYIIPGSTNPEYFKEDK 300
 QY 301 QPQNMAGNFMNLSQGCDDQASPGILRLQFVLYOHFQPNESNRIYVVDLSNERAMSLTIE 360
 DB 301 QPQNMAGNFMNLSQGCDDQASPGILRLQFVLYOHFQPNESNRIYVVDLSNERAMSLTIE 360
 QY 361 PRVYKQSRKRVPCGFCVLESRTCSNLTLSGSKHKISFLCDLTRLMMNVEKTICTDH 420
 DB 361 PRVYKQSRKRVPCGFCVLESRTCSNLTLSGSKHKISFLCDLTRLMMNVEKTICTDH 420
 QY 421 RYQQRKSYSLQVPSDLHLHPVELHDFSWMKLVPRKDRSLVLPAPKQOQHTHKPCQTSF 480
 DB 421 RYQQRKSYSLQVPSDLHLHPVELHDFSWMKLVPRKDRSLVLPAPKQOQHTHKPCQTSF 480
 QY 481 SYLVASAIPIPSQDLFYSGFCGSGSIKOIYQKNTSVTLTRFAPSFQOASNOGLTVSITPY 540
 DB 481 SYLVASAIPIPSQDLFYSGFCGSGSIKOIYQKNTSVTLTRFAPSFQOASNOGLTVSITPY 540
 QY 541 FKEEGYFTVPTDTKSKYVLTLPWMDRGLPSLVSNVISVPRQVACLFFKERSGVYCC 600
 DB 541 FKEEGYFTVPTDTKSKYVLTLPWMDRGLPSLVSNVISVPRQVACLFFKERSGVYCC 600
 QY 601 TGAAPMIOQORTRAEITFLDDEDVLPKPSFHHHSFVWNISNCSPTSGKDLDFSTLT 660
 DB 601 TGAAPMIOQORTRAEITFLDDEDVLPKPSFHHHSFVWNISNCSPTSGKDLDFSTLT 660
 QY 661 PRVVDLTVLIIAIVGGVLLLSALGLITICVKKKKKTKNGPVGITNGINTEMPRQK 720
 DB 661 PRVVDLTVLIIAIVGGVLLLSALGLITICVKKKKKTKNGPVGITNGINTEMPRQK 720
 QY 721 KFOGGRKNDNSHVAVIEDTMVYGHLLQDSSGSFLQPEVDIYRPFQGTGVCPPSPPTIC 780

DB 721 KFOGGRKNDNSHVAVIEDTMVYGHLLQDSSGSFLQPEVDIYRPFQGTGVCPPSPPTIC 780
 QY 781 SRAPTAKLATEEPPRPSPSESEBPYTFSSHNNQDVSSKRTDIPILNTQEMPEPAE 836
 DB 781 SRAPTAKLATEEPPRPSPSESEBPYTFSSHNNQDVSSKRTDIPILNTQEMPEPAE 836
 RESULT 2
 AAY91456
 ID AAY91456 standard; Protein; 709 AA.
 XX
 AC AAY91456;
 XX
 DT 29-JUN-2000 (first entry)
 XX
 DE Human secreted protein sequence encoded by gene 6 SEQ ID NO:129.
 XX
 KW Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
 KW antiHIV; antiinflammatory; nootropic; neuroprotective; antiallergic;
 KW osteopathic; antiarthritic; antibacterial; antidiabetic; antisthma;
 KW immunosuppressive; antiHIV; antiinflammatory; nootropic; neuroprotective;
 KW antiallergic; osteopathic; antiarthritic; antibacterial; antidiabetic;
 KW antisthma; antipsoatic; and cardiant. The polynucleotides and their
 CC corresponding secreted proteins are useful for preventing, treating or
 CC ameliorating medical conditions, e.g. by protein or gene therapy. Also
 CC pathological conditions can be diagnosed by determining the amount of the
 CC proteins in a sample or by determining the presence of mutations in the
 CC polynucleotides. Specific uses are described for each of the
 CC polynucleotides, based on which tissues they are most highly expressed
 CC in, and include developing products for the diagnosis or treatment of
 CC cancer, tumours, neurodegenerative disorders, developmental abnormalities
 CC and foetal deficiencies, blood disorders, diseases of the immune system,
 CC autoimmune diseases, hepatic and renal disease, inflammation,
 CC allergies, Alzheimer's and behavioural disorders, schizophrenia,
 CC osteoporosis, arthritis, infections, AIDS, spinal cord injuries,
 CC transplant rejection, diabetes, asthma, sepsis, acne, psoriasis,
 CC
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Komatsu, Iis GA, Rosen CA, Ruben SM, Duan R, Moore PA, Shi Y;
 PI Lafleur D, Wei Y, Ni J, Florence KA, Young PE, Brewer LA;
 PI Soppet DR, Endress CA, Ebner R, Olsen HS, Mucenski M;
 XX
 DR WPI: 2000-195282/17.
 XX N-P-SDB: AAA26351.
 PT New isolated human genes and the secreted polypeptides they encode,
 PT useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, inflammation or blood disorders -
 XX
 PS Claim 11; Page 456-459; 634pp; English.
 XX
 CC The polynucleotide sequences given in AAA26346 to AAA26458 encode the
 CC human secreted proteins given in AAY91451 to AAY91691. The human secreted
 CC proteins can have activities based on the tissues and cells they are
 CC expressed in. Examples of the activities are: cytostatic;
 CC immunosuppressive; antiHIV; antiinflammatory; nootropic; neuroprotective;
 CC antiallergic; osteopathic; antiarthritic; antibacterial; antidiabetic;
 CC antisthma; antipsoatic; and cardiant. The polynucleotides and their
 CC corresponding secreted proteins are useful for preventing, treating or
 CC ameliorating medical conditions, e.g. by protein or gene therapy. Also
 CC pathological conditions can be diagnosed by determining the amount of the
 CC proteins in a sample or by determining the presence of mutations in the
 CC polynucleotides. Specific uses are described for each of the
 CC polynucleotides, based on which tissues they are most highly expressed
 CC in, and include developing products for the diagnosis or treatment of
 CC cancer, tumours, neurodegenerative disorders, developmental abnormalities
 CC and foetal deficiencies, blood disorders, diseases of the immune system,
 CC autoimmune diseases, hepatic and renal disease, inflammation,
 CC allergies, Alzheimer's and behavioural disorders, schizophrenia,
 CC osteoporosis, arthritis, infections, AIDS, spinal cord injuries,
 CC transplant rejection, diabetes, asthma, sepsis, acne, psoriasis,

cardiovascular disorders, reproductive disorders, gastrointestinal disorders, respiratory disorders and metabolic disorders. The proteins or polynucleotides can also be used as food additives or preservatives. The proteins are also useful for identifying their binding partners. CC AA26337 to AA26345 and AA291450 are sequences used in the CC exemplification of the present invention.

Sequence 709 AA:

Query Match 82.7%; Score 3634.5; DB 21; Length 709;
Best Local Similarity 98.5%; Pred. No. 6.1e-308;
Matches 699; Conservative 2; Mismatches 8; Indels 1; Gaps 1;

```

OY 1 MAGLNGVSIALGVLLGAARLPKGAFAETALPRESNTYLKIGTPTLLAKPCYIYI 60
DB 1 MAGLNGVSIALGVLLGAARLPKGAFAETALPRESNTYLKIGTPTLLAKPCYIYI 60
OY 61 SKRHITMLSIKSGERIVTFSCQSPENHVEIQRNIDMSGPCPEGEVQLPSTSLPT 120
DB 61 SKRHITMLSIKSGERIVTFSCQSPENHVEIQRNIDMSGPCPEGEVQLPSTSLPT 120
OY 121 LMRITMDVKAKHSIGLELOFSIPRLRQIGPESCPDGYTHSISGRIDATVVRIGTCSN 180
DB 121 LMRITMDVKAKHSIGLELOFSIPRLRQIGPESCPDGYTHSISGRIDATVVRIGTCSN 180
OY 181 GYVSRTKKQEGVMALHLPWFHPRNVSIGSIANRSSIKRLCTIESEVEEGSNTILMSAN 240
DB 181 GYVSRTKKQEGVMALHLPWFHPRNVSIGSIANRSSIKRLCTIESEVEEGSNTILMSAN 240
OY 241 PGCFPEDELMTQFVPAHLRASVSFLNENLSNCEKREVEYIYIGSTINPEVFLKEDK 300
DB 241 PGCFPEDELMTQFVPAHLRASVSFLNENLSNCEKREVEYIYIGSTINPEVFLKEDK 300
OY 301 QPGNAGNENLSLQGCDDQDQSPGILRLQFVLYOHQPNESKRIYVVDLSNERAMSLTE 360
DB 301 QPGNAGNENLSLQGCDDQDQSPGILRLQFVLYOHQPNESKRIYVVDLSNERAMSLTE 360
OY 361 PPRVQSRKFEVPGCFVCSRTSSNLTLTSGSKHISFLCDLRLMANNVEKTTSCMTH 420
DB 361 PPRVQSRKFEVPGCFVCSRTSSNLTLTSGSKHISFLCDLRLMANNVEKTTSCMTH 420
OY 421 RYCQRKSYSLQVPSDILHLPVELHDFSMKLVPRKDLISLVLPKQLOOHTHEKPCNTSF 480
DB 421 RYCQRKSYSLQVPSDILHLPVELHDFSMKLVPRKDLISLVLPKQLOOHTHEKPCNTSF 480
OY 481 SYLVASAIQSDLYFGSGFCGSGIKQIQVKNISVTLRTFAPSFOQEASRQGLTVSFIY 540
DB 481 SYLVASAIQSDLYFGSGFCGSGIKQIQVKNISVTLRTFAPSFOQEASRQGLTVSFIY 540
OY 541 FKEEGVFTVPTPKSKVYLRTPMNDGLSLTSSVSNISVPRDQVACLFFEKERSVQO 600
DB 541 FKEEGVFTVPTPKSKVYLRTPMNDGLSLTSSVSNISVPRDQVACLFFEKERSVQO 600
OY 601 TGRARMIIOEQRTAEIISLDEDLPRKPSFHHSFVWNISCSPTSGKQLDLFFSVTLT 660
DB 601 TGRARMIIOEQRTAEIISLDEDLPRKPSFHHSFVWNISCSPTSGKQLDLFFSVTLT 660
OY 661 PRTVDLTVLIAVGGVLLSALGLIICVKKKKTKNGKPAVGIYNGN 710
DB 661 PRTVDLTVLIAVGGVLLSALGLIICVKKKKTKNGKPAVGIYNGN 710

```

RESULT 3

AA26340 standard; Protein; 749 AA.

AA26340;

17-MAY-2002 (first entry)

Human tumour-associated antigen B345 protein.

Tumour-associated antigen; human; B345; cytosolic; cell communication;

cell interaction; signal transduction; metastasis; cancer; colon; immunotherapy; carcinoma; lung; diagnosis.

Homo sapiens.

WO200204508-A1.

17-JAN-2002.

05-JUL-2001; 2001WO-BP07705.

07-JUL-2000; 2000DE-1033080.

19-APR-2001; 2001DE-1019294.

(BOEH) BOEHRINGER INGELHEIM INT GMBH.

Schweifer N, Scherl-Mostageer M, Sommergruber W, Abseher R;

WPI: 2002-171704/22.

N-PSDB: ABA99506.

New tumour-associated antigen B345, useful for diagnosis and immunotherapy of tumors, also related nucleic acid and antibodies

Example 6; Page 76-78; 102pp; German.

This invention describes a novel tumour-associated antigen, designated B345 which has cytostatic activity. B345 is involved in communication, interaction and/or signal transduction with extracellular components and ligands, especially in the metastatic potential of cancers, particularly of the colon. B345 or its immunogenic fragments, also the DNA that encodes it, are useful for immunotherapy of cancer, particularly carcinoma of lung or colon. Antibodies raised against B345 are useful for treatment and diagnosis of cancers that are associated with B345 expression, including their use for targeted delivery of cytotoxic or radioactive agents. Probes derived from B345 can be used to detect tumour-specific mutations in the B345 sequence, and can be used to screen for B345 specific modulators. This sequence represents a human B345 tumour-associated antigen described in the invention.

Sequence 749 AA:

Query Match 78.9%; Score 3467; DB 23; Length 749;
Best Local Similarity 90.3%; Pred. No. 2.8e-293;
Matches 676; Conservative 9; Mismatches 28; Indels 36; Gaps 6;

```

OY 67 MLISKGERIVTFSCQSPENHVEIQRNIDMSGPCPEGEVQLPSTSLPTLNRPTI 126
DB 1 MLISKGERIVTFSCQSPENHVEIQRNIDMSGPCPEGEVQLPSTSLPTLNRPTI 126
OY 127 MDYKAKHSIGLELOFSIPRLRQIGPESCPDGYTHSISGRIDATVVRIGTCSNGTVSRN 186
DB 127 MDYKAKHSIGLELOFSIPRLRQIGPESCPDGYTHSISGRIDATVVRIGTCSNGTVSRN 186
OY 187 KMOEGVMALHLPWFHPRNVSIGSIANRSSIKRLCTIESEVEEGSNTILMSANYPGPE 246
DB 187 KMOEGVMALHLPWFHPRNVSIGSIANRSSIKRLCTIESEVEEGSNTILMSANYPGPE 246
OY 247 DELMTQFVPAHLRASVSFLNENLSNCEKREVEYIYIGSTINPEVFLKEDKQGNNA 306
DB 247 DELMTQFVPAHLRASVSFLNENLSNCEKREVEYIYIGSTINPEVFLKEDKQGNNA 306
OY 307 GNFNLSLQGCDDQDQSPGILRLQFVLYOHQPNESKRIYVVDLSNERAMSLTEPRVQO 366
DB 307 GNFNLSLQGCDDQDQSPGILRLQFVLYOHQPNESKRIYVVDLSNERAMSLTEPRVQO 366
OY 367 SRKFPVPGCFVCSRTSSNLTLTSGSKHISFLCDLRLMANNVEKTTSCDHRKQCK 426
DB 367 SRKFPVPGCFVCSRTSSNLTLTSGSKHISFLCDLRLMANNVEKTTSCDHRKQCK 426
OY 427 SYSLQVPSDILHLPVELHDFSMKLVPRKDLISLVLPKQLOOHTHEKPCNTSFSVLVS 486
DB 427 SYSLQVPSDILHLPVELHDFSMKLVPRKDLISLVLPKQLOOHTHEKPCNTSFSVLVS 486

```


D6	30	MAQNCVSTALILGVLLGLAARLRPGAEAFETLAPRESNITVLINLKGPPTTLAKPCIVY	89
OY	61	SKRHITMLSIKSGRIYVFTEFSCSPENHVEIEIQNKIDCMGBCPGEVOLOPSTSLPT	120
D6	90	SKRHITMLSIKSGRIYVFTEFSCSPENHVEIEIQNKIDCMGBCPGEVOLOPSTSLPT	149
OY	121	LNRFTIDVAHNSIGLEIOFSIPRLROTGPGESCDDGTHTSISGRIDATVYRIGTFCSN	180
D6	150	LNRFTIDVAHNSIGLEIOFSIPRLROTGPGESCDDGTHTSISGRIDATVYRIGTFCSN	209
OY	181	GTVSRIMOGEGVMALLLPWFHPHRNVSGFSIARRSSIKRLCTIESEFEESGATLMASNY	240
D6	210	GTVSRIMOGEGVMALLLPWFHPHRNVSGFSIARRSSIKRLCTIESEFEESGATLMASNY	269
OY	241	PEGFEDELMTWOFVVPAPHLRASVSFLNFNLSCERKEEREVEYYIPGSTTNEVEFKLEDK	300
D6	270	PEGFEDELMTWOFVVPAPHLRASVSFLNFNLSCERKEEREVEYYIPGSTTNEVEFKLEDK	329
OY	301	QPGMANFNMLSLGGCDQDQOSPGILRLQFOYLVOHPQESNKIYVVDLSNRAMSLTE	360
D6	330	QPGMANFNMLSLGGCDQDQOSPGILRLQFOYLVOHPQESNKIYVVDLSNRAMSLTE	389
OY	361	PRPVKGRRKRVPGCFVCLESRTSSNLTLTSGSKHKISFLCDDLPLTMANNVK	413
D6	390	PRPVKGRRKRVPGCFVCLESRTSSNLTLTSGSKHKISFLCDDLPLTMANNVK	442
<hr/>			
RESULT 6			
ID	AAU12252	standard; protein; 343 AA.	
AC	AAU12252:		
XX	24-OCT-2001	(first entry)	
DE	Human PRO5773	polypeptide sequence.	
OS	Homo sapiens.		
XX	MO200140466-A2.		
XX	07-JUN-2001.		
PE	01-DEC-2000;	2000MO-US32678.	
XX	01-DEC-1999;	99MO-US28301.	
XX	01-DEC-1999;	99MO-US28634.	
XX	02-DEC-1999;	99MO-US28551.	
XX	02-DEC-1999;	99MO-US28564.	
XX	02-DEC-1999;	99MO-US28565.	
XX	09-DEC-1999;	99US-0170262.	
XX	16-DEC-1999;	99MO-US30095.	
XX	20-DEC-1999;	99MO-US30911.	
XX	30-DEC-1999;	99MO-US30999.	
XX	30-DEC-1999;	99MO-US31243.	
XX	06-JAN-2000;	2000MO-US00377.	
XX	06-JAN-2000;	2000MO-US00376.	
XX	11-FEB-2000;	2000MO-US03565.	
XX	18-FEB-2000;	2000MO-US04341.	
XX	18-FEB-2000;	2000MO-US04342.	
XX	22-FEB-2000;	2000MO-US04414.	
XX	24-FEB-2000;	2000MO-US04914.	
XX	24-FEB-2000;	2000MO-US05004.	
XX	01-MAR-2000;	2000MO-US05601.	
XX	20-MAR-2000;	2000MO-US07377.	
XX	30-MAR-2000;	2000MO-US07532.	
XX	17-MAY-2000;	2000MO-US13705.	

22-MAY-2000: 2000MO-US14042.
 PR 30-MAY-2000: 2000MO-US14941.
 PR 02-JUN-2000: 2000MO-US15264.
 PR 10-NOV-2000: 2000MO-US30873.
 XX
 PA (GETH) GENENTECH INC.
 XX Baker KP, Beresini M, DeForge L, Desnoyers L, Filvaroff E, Gao W;
 PI Gerlitsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
 PI Smith V, Stewart TA, Tamas D, Watanabe CK, Wood WI, Zhang Z;
 XX
 DR WPI: 2001-408281/43.
 DR N-PSDB; AAS21324.
 XX
 PT Isolated, secretory and transmembrane PRO polypeptide used to detect
 PT other PRO polypeptides, link bioactive molecules to cells expressing
 PT PRO polypeptides, and detect the presence of mammalian tumours e.g.
 PT lung, breast, prostate, cervical -
 XX
 PS Claim 12; Fig 162; 813pp; English.
 XX
 CC AAU12172-AAU12446 represent novel human secretory and transmembrane
 CC PRO polypeptides. The PRO polypeptides are useful to detect other
 CC PRO polypeptides, to link bioactive molecules to cells expressing
 CC PRO polypeptides, to modulate biological activities of cells expressing
 CC PRO polypeptides, and to detect the presence of mammalian lung, colon,
 CC breast, prostate, rectal, cervical or liver tumours by comparing PRO
 CC polypeptide expression in a cell sample to that in a control sample.
 CC Some of the 275 sequences are also useful to stimulate the release of
 CC tumour necrosis factor-alpha (TNF-alpha) from human blood, the
 CC proliferation or differentiation of chondrocytes, the proliferation or
 CC gene expression in pericyte cells, the release of proteoglycans from
 CC cartilage, the proliferation of inner ear utricular supporting cells or
 CC of T-lymphocytes, the release of a cytokine from peripheral blood
 CC monocytes (PBMCs), or the proliferation of endothelial cells. Some of
 CC the PRO polypeptides may modulate glucose or free fatty acid uptake by
 CC skeletal muscle cells or by adipocytes, or inhibit binding of A-peptide
 CC to factor VIIA. The PRO polypeptides can be used in assays to identify
 CC molecules involved in binding interactions. The polynucleotides encoding
 CC PRO polypeptides can be used to generate probes, antisense RNA/DNA,
 CC transgenic or knock out animals and can be used in gene therapy.
 CC
 XX
 SO Sequence 343 AA;

Query Match 40.5%; Score 1779; DB 22; Length 343;
 Best Local Similarity 99.4%; Pred. No. 1.8e-146;
 Matches 341; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGNGCVSALLGVLLGAAALPRGAEFIALPRESNTVILKICTPTLLAKPCYIV 60
 DB 1 MAGNGCVSALLGVLLGAAALPRGAEFIALPRESNTVILKICTPTLLAKPCYIV 60
 QY 61 SKRHITMLSIKSGERTVETFSQSPENHVEIEIQKNIDCSGPCPEGEVOLQPSLPT 120
 DB 61 SKRHITMLSIKSGERTVETFSQSPENHVEIEIQKNIDCSGPCPEGEVOLQPSLPT 120
 QY 121 LNTFTMDVKAHKSIGLELOFSTPLRLQIGPSCDGVTHSISGRIDATVVRIGTFCSN 180
 DB 121 LNTFTMDVKAHKSIGLELOFSTPLRLQIGPSCDGVTHSISGRIDATVVRIGTFCSN 180
 QY 181 GTSRIRKMGQVKAHLHPFHRNVSQFSIARSSIKRCLCTIESVEEGSGATLMSAN 240
 DB 181 GTSRIRKMGQVKAHLHPFHRNVSQFSIARSSIKRCLCTIESVEEGSGATLMSAN 240
 QY 241 PEGFPEDELMTQFVPAHLRASVSFLNPLNSCERKEEVEYITGSTNPNPFKLEDK 300
 DB 241 PEGFPEDELMTQFVPAHLRASVSFLNPLNSCERKEEVEYITGSTNPNPFKLEDK 300
 QY 301 QPQNMAGNFNLSLQCGDQDQSPGILRLQPVLYVQHPRQNSNK 343
 DB 301 QPQNMAGNFNLSLQCGDQDQSPGILRLQPVLYVQHPRQNSNK 343

RESULT 7
 ID AAY91593
 XX AAY91593 standard; Protein: 159 AA.
 AC AAY91593;
 XX
 DT 29-JUN-2000 (first entry)
 XX
 DE Human secreted protein sequence encoded by gene 6 SEQ ID NO:266.
 XX
 KW Human: secreted protein; diagnosis; cytostatic; immunosuppressive;
 KW antiHIV; antiinflammatory; neutropic; neuroprotective; antiallergic;
 KW osteopathic; antidiabetic; antibacterial; antidiabetic; antisthma;
 KW antiparasitic; cardiac; gene therapy; cancer; neurological disorder;
 KW immune disease; inflammation; blood disorder; tumour.
 XX
 OS Homo sapiens.
 XX
 PN WO200006698-A1.
 PD 10-FEB-2000.
 XX
 PF 29-JUL-1999; 99MO-US17130.
 XX
 PR 30-JUL-1998; 98US-0094657.
 PR 05-AUG-1998; 98US-0095486.
 PR 06-AUG-1998; 98US-0095454.
 PR 06-AUG-1998; 98US-0095455.
 PR 12-AUG-1998; 98US-0096319.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Komatsoulis GA, Rosen CA, Ruben SM, Duan R, Moore PA, Shi Y;
 PI Lafleur D, Wei Y, Ni J, Florence KA, Young PE, Brewer LA;
 PI Soppet DR, Endress GA, Ebner R, Olsen HS, Mucenksi M;
 XX
 DR WPI: 2000-195282/17.
 XX
 PT New isolated human genes and the secreted polypeptides they encode,
 PT useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, inflammation or blood disorders -
 XX
 PS Disclosure: Page 18; 634pp; English.

The polynucleotide sequences given in AAA26346 to AAA26458 encode the
 CC human secreted proteins given in AAY91451 to AAY91691. The human secreted
 CC proteins can have activities based on the tissues and cells they are
 CC expressed in. Examples of the activities are: cytostatic;
 CC immunosuppressive; antiHIV; antiinflammatory; neutropic; neuroprotective;
 CC antiallergic; osteopathic; antidiabetic; antibacterial; antidiabetic;
 CC antisthma; antiparasitic; and cardiac. The polynucleotides and their
 CC corresponding secreted proteins are useful for preventing, treating or
 CC ameliorating medical conditions, e.g. by protein or gene therapy. Also
 CC pathological conditions can be diagnosed by determining the amount of the
 CC proteins in a sample or by determining the presence of mutations in the
 CC polynucleotides. Specific uses are described for each of the
 CC polynucleotides, based on which tissues they are most highly expressed
 CC in, and include developing products for the diagnosis or treatment of
 CC cancer, tumours, neurodegenerative disorders, developmental abnormalities
 CC and foetal deficiencies, blood disorders, diseases of the immune system,
 CC autoimmune diseases, hepatic and renal disease, inflammation,
 CC allergies, Alzheimer's and behavioural disorders, schizophrenia,
 CC osteoporosis, arthritis, infections, AIDS, spinal cord injuries,
 CC transplant rejection, diabetes, asthma, sepsis, acne, psoriasis,
 CC cardiovascular disorders, reproductive disorders, gastrointestinal
 CC disorders, respiratory disorders and metabolic disorders. The proteins
 CC or polynucleotides can also be used as food additives or preservatives.
 CC The proteins are also useful for identifying their binding partners.
 CC AAA26337 to AAA26345 and AAY91450 are sequences used in the
 CC exemplification of the present invention.
 XX
 SO Sequence 159 AA;

Query Match 18.8%; Score 827; DB 21; Length 159;
 Best Local Similarity 100.0%; Pred. No. 6-9e-64;
 Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 FEIALPRESNITVLKLGPTLLAKPCYIVISKRHITMISIKSGERIVTFSCOSPENHF 89
 DB 1 FEIALPRESNITVLKLGPTLLAKPCYIVISKRHITMISIKSGERIVTFSCOSPENHF 60
 QY 90 VIEIQNNICMCGPCFGEVQLOPSTSLPTLTKRTFIMDVKAHKSIGLEQESIPRLRQI 149
 DB 61 VIEIQNNICMCGPCFGEVQLOPSTSLPTLTKRTFIMDVKAHKSIGLEQESIPRLRQI 120
 QY 150 GPGEPCPDGVTHSISGRIDATVVRIGTFCSNGTVSRIMK 188
 DB 121 GPGEPCPDGVTHSISGRIDATVVRIGTFCSNGTVSRIMK 159

RESULT 8
 ABG00296
 ID ABG00296 standard; Protein: 143 AA.
 AC ABG00296;
 DT 13-FEB-2002 (first entry)
 DE Novel human diagnostic protein #287.
 XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX Homo sapiens.
 OS WO200175067-A2.
 PN 11-OCT-2001.
 PD 30-MAR-2001; 2001WO-US08631.
 PF 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX (HYSE-) HYSEQ INC.
 PA Drmanac RT, Liu C, Tang YF;
 PI WPI: 2001-639362/73.
 DR N-PSDB; AAS64483.
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 PT Claim 20; SEQ ID No 30655; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 143 AA;
 QY Query Match 17.6%; Score 775; DB 22; Length 143;
 DB Best Local Similarity 100.0%; Pred. No. 2e-59; Indels 0; Gaps 0;
 CC Matches 143; Conservative 0; Mismatches 0;

QY 694 KKKTKNGPAVGIVNGNINTEMPROPKRFQGRKNDSHYVAVIEDTMVYGHLLDSSGS 753
 DB 1 KKKTKNGPAVGIVNGNINTEMPROPKRFQGRKNDSHYVAVIEDTMVYGHLLDSSGS 60
 QY 754 FLOPEVDYTRPFQGTMGVCPSPPTICRAPTAKLATEEPPSPSESEPYTFSHPN 813
 DB 61 FLOPEVDYTRPFQGTMGVCPSPPTICRAPTAKLATEEPPSPSESEPYTFSHPN 120
 QY 814 GDVSSKDTDIPLLNTOEPMEPAE 836
 DB 121 GDVSSKDTDIPLLNTOEPMEPAE 143

RESULT 9
 AAM23969
 ID AAM23969 standard; Protein: 121 AA.
 AC AAM23969;
 DT 12-OCT-2001 (first entry)
 DE Human EST encoded protein SEQ ID NO: 1494.
 XX Human: sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
 KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
 KW diagnostics; forensic test; gene mapping; genetic disorder;
 XX biodiversity; gene therapy; nutrition.
 OS Homo sapiens.
 PN WO200154477-A2.
 PD 02-AUG-2001.
 PE 25-JAN-2001; 2001WO-US02687.
 PF 25-JAN-2000; 2000US-0491404.
 PR 17-JUL-2000; 2000US-0617746.
 PR 03-AUG-2000; 2000US-0631451.
 PR 15-SEP-2000; 2000US-0663870.
 XX (HYSE-) HYSEQ INC.
 PA Tang YF, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
 PI Cao Y, Drmanac RA, Zhang J, Werhman T;
 DR WPI: 2001-476164/51.
 XX N-PSDB; AAH98628.
 XX Isolated polypeptide for treatment of diseases, diagnostics, raising
 PT antibodies and research use -
 PT Claim 20; Page 1036; 1275pp; English.

CC The present invention provides the protein and coding sequences of novel
 CC proteins from a variety of organisms, including human, dog, cat, horse,
 CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
 CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
 CC from the organism of interest. They can be used in diagnostics,
 CC forensics, gene mapping, identification of mutations, to assess
 CC biodiversity and for nutritional purposes. The present sequence is a
 CC protein of the invention.

Sequence 121 AA;
Query Match 14.0%; Score 616; DB 22; Length 121;
Best Local Similarity 100.0%; Pred. No. 1.2e-45;
Matches 121; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGINCVSIALGCVLLGAAARLPRGAEAFIALPRESNTIVLKIGTPILLARPCYIYI 60
D 1 MAGINCVSIALGCVLLGAAARLPRGAEAFIALPRESNTIVLKIGTPILLARPCYIYI 60
QY 61 SKRIITMSTKSGERIVFTFSCSPENHFVIEIKNIDCMGPGCPGEVOLQPSSTLPT 120
D 61 SKRIITMSTKSGERIVFTFSCSPENHFVIEIKNIDCMGPGCPGEVOLQPSSTLPT 120
QY 121 L 121
D 121 L 121
Db 121 L 121

RESULT 10
AAU69483
ID AAU69483 standard; Protein: 167 AA.

AC AAU69483;
XX
DT 29-JAN-2002 (first entry)
XX
DE Human purified secretory polypeptide #52.
XX
KW Human; purified secretory polypeptide; cell proliferative disorder;
KW cancer; immune system disorder; neurological disorder; mental disorder;
KW motor neuron disorder; demyelinating disorder; neuromuscular disorder;
KW central nervous system disorder; enzyme linked immunosorbent assay;
KW ELISA; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200162918-A2.
XX
PD 30-AUG-2001.
XX
PE 01-FEB-2001; 2001MO-US03465.
XX
PR 24-FEB-2000; 2000US-185215P.
PR 24-FEB-2000; 2000US-185216P.
PR 16-MAY-2000; 2000US-205232P.
PR 17-MAY-2000; 2000US-205286P.
PR 17-MAY-2000; 2000US-205287P.
PR 17-MAY-2000; 2000US-205323P.
PR 17-MAY-2000; 2000US-205324P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Panzer SR, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC;
PI Chen A, D'sa SA, Amshay S, Dahl CR, Dam JC, Daniels SE;
PI Dufour GE, Flores V, Fong WT, Greenawalt IB, Hillman JL, Jones AL;
PI Liu TF, Roseberry AM, Rosen BH, Russo FD, Stockdreher TK, Daffo A;
PI Wright RJ, Yap PE, Yu JY, Bradley DL, Bratcher SR, Chen W;
PI Cohen HJ, Hodgson DM, Lincoln SE;
XX
DR WPI; 2001-648217/4.
XX
XX Nucleic acids encoding secretory polypeptides, useful in genetic
PT diagnosis and therapy -
XX
PS Disclosure; Page 221; 237pp; English.

CC Sequences AAU69432-AAU69511 represent purified secretory polypeptides of
CC the invention. The polypeptides and their associated polynucleotides can
CC be used in the treatment, prevention and diagnosis of diseases associated
CC with inappropriate secretory protein expression. These diseases include
CC proliferative disorders such as atherosclerosis and psoriasis,
CC as well as leukæmia and melanoma, immune system disorders such as

CC asthma and diabetes mellitus, neurological disorders such as epilepsy and
CC Parkinson's disease, mental disorders such as schizophrenia and seasonal
CC affective disorder (SAD), motor neuron disorders such as amyotrophic
CC lateral sclerosis, demyelinating disorders such as multiple sclerosis,
CC central nervous system disorders such as mental retardation and
CC neurofibromatosis and neuromuscular disorders such as cerebral palsy and
CC muscular dystrophy. Target polynucleotides in a sample can be detected by
CC hybridising the sample with a probe sequence complementary to the target
CC polynucleotide, under conditions in which a hybridisation complex is
CC formed, and detecting the presence or absence of the complex. The
CC polypeptides may also be used as antigens in the production of antibodies
CC against secretory proteins and in assays to identify modulators of
CC protein expression and activity. The antibodies may also be used as
CC diagnostic agents for detecting the presence of the sequences of the
CC invention in samples e.g. by enzyme linked immunosorbent assay (ELISA).
XX

Sequence 167 AA;
Query Match 11.9%; Score 523; DB 22; Length 167;
Best Local Similarity 73.4%; Pred. No. 2.6e-37;
Matches 113; Conservative 7; Mismatches 32; Indels 2; Gaps 2;

QY 94 OKNIDCMGPGCP-EGEVOLOPSTSLPLNTFTMDVKAHKSIGLELOFSIPRLROIGPG 152
D 9 RKNIELYRPMSLGEGVOLQPSLPLNTFTMDVKAHKSIGLELOFSIPRLROIGPV 68
QY 153 ESCPDGYTHSISGRIDATVVRIGTSCNGTVSRIRKMGCVKALHPV-FHPRNVSGFSI 211
D 69 KSCPDDGYTHSISGRIDATVVRIGTSCNGTVSRIRKMGCVKALHPV-FHPRNVSGFSI 128
QY 212 ANRSSIRKLCTIEVFEGEGSATLMSANYPEGFP 245
D 129 ETRSSIRKLCTIEVFEGEGSATLNVCOLPYKAP 162

RESULT 11
AAM90342
ID AAM90342 standard; Protein: 116 AA.
XX
AC AAM90342;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen SEQ ID NO:17935.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PE 17-JAN-2001; 2001MO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 31-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.

```
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225477.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227189.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0233223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0235837.
PR 29-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.

PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251088.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0251679.
PR 06-DEC-2000; 2000US-0251679.
PR 06-DEC-2000; 2000US-0251679.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
PA
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX N-PSDB; AAK63123.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS
XX
XX Claim 11: SEQ ID NO 17935; 3071p + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAK62170 to AAK61921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
```

CC Cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX

SQ Sequence 116 AA;

Query Match 8.9%; Score 390; DB 22; Length 116;
Best Local Similarity 98.7%; Pred. No. 5.9e-26;
Matches 75; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 341 SNKTYVVDLSNERAMSLTTPRPVKSRKFPVCGFCVLESRTSSNLTITSGSKHKRISFL 400
|:|||||
Db 12 SDRITYVDLSNERAMSLTTPRPVKSRKFPVCGFCVLESRTSSNLTITSGSKHKRISFL 71
|:|||||

OY 401 CDDLTRLMNVKRTIS 416
|:|||||
Db 72 CDDLTRLMNVKRTIS 87
|:|||||

RESULT 12
AAV27020
ID AAV27020 standard; Protein; 3623 AA.
XX AAV27020;

XX 08-OCT-1999 (first entry)

XX Rat cubilin protein sequence.

XX Cubilin; epithelial glycoprotein receptor; vitamin B12; endocytosis;
XX toxicity; kidney; spleen; brain; liver; heart; thyroid; abortion; rat;
XX proteinuria; fetal malformation; fetal development; kidney damage.

XX Rattus sp.

XX Key Location/Qualifiers

FT Peptide 1..20 /note- "endoplasmatic import signal sequence"

FT Protein 20 /note- "mature protein"

FT Modified-site 95 /note- "potential N-glycosylation site"

FT Modified-site 428 /note- "potential N-glycosylation site"

FT Modified-site 491 /note- "potential N-glycosylation site"

FT Modified-site 711 /note- "potential N-glycosylation site"

FT Modified-site 781 /note- "potential N-glycosylation site"

FT Modified-site 857 /note- "potential N-glycosylation site"

FT Modified-site 957 /note- "potential N-glycosylation site"

FT Modified-site 984 /note- "potential N-glycosylation site"

FT Modified-site 1168 /note- "potential N-glycosylation site"

FT Modified-site 1285 /note- "potential N-glycosylation site"

FT Modified-site 1307 /note- "potential N-glycosylation site"

FT Modified-site 1319 /note- "potential N-glycosylation site"

FT Modified-site 1332 /note- "potential N-glycosylation site"

FT Modified-site 1500 /note- "potential N-glycosylation site"

FT Modified-site 1551 /note- "potential N-glycosylation site"

FT Modified-site 1646 /note- "potential N-glycosylation site"

FT Modified-site /note- "potential N-glycosylation site"

FT Modified-site 1671 /note- "potential N-glycosylation site"
FT Modified-site 1802 /note- "potential N-glycosylation site"
FT Modified-site 1819 /note- "potential N-glycosylation site"
FT Modified-site 2085 /note- "potential N-glycosylation site"
FT Modified-site /note- "potential N-glycosylation site"
FT Modified-site 2117 /note- "potential N-glycosylation site"
FT Modified-site 2274 /note- "potential N-glycosylation site"
FT Modified-site 2400 /note- "potential N-glycosylation site"
FT Modified-site /note- "potential N-glycosylation site"
FT Modified-site 2531 /note- "potential N-glycosylation site"
FT Modified-site 2581 /note- "potential N-glycosylation site"
FT Modified-site /note- "potential N-glycosylation site"
FT Modified-site 2610 /note- "potential N-glycosylation site"
FT Modified-site 2813 /note- "potential N-glycosylation site"
FT Modified-site 2875 /note- "potential N-glycosylation site"
FT Modified-site 2945 /note- "potential N-glycosylation site"
FT Modified-site /note- "potential N-glycosylation site"
FT Modified-site 2989 /note- "potential N-glycosylation site"
FT Modified-site /note- "potential N-glycosylation site"
FT Modified-site 3042 /note- "potential N-glycosylation site"
FT Modified-site /note- "potential N-glycosylation site"
FT Modified-site 3106 /note- "potential N-glycosylation site"
FT Modified-site 3125 /note- "potential N-glycosylation site"
FT Modified-site /note- "potential N-glycosylation site"
FT Modified-site 3165 /note- "potential N-glycosylation site"
FT Modified-site /note- "potential N-glycosylation site"
FT Modified-site 3288 /note- "potential N-glycosylation site"
FT Modified-site 3283 /note- "potential N-glycosylation site"
FT Modified-site /note- "potential N-glycosylation site"
FT Modified-site 3290 /note- "potential N-glycosylation site"
FT Modified-site /note- "potential N-glycosylation site"
FT Modified-site 3357 /note- "potential N-glycosylation site"
FT Modified-site /note- "potential N-glycosylation site"
FT Modified-site 3400 /note- "potential N-glycosylation site"
FT Modified-site 3430 /note- "potential N-glycosylation site"
FT Modified-site /note- "potential N-glycosylation site"
FT Modified-site 3533 /note- "potential N-glycosylation site"
FT Modified-site /note- "potential N-glycosylation site"
FT Modified-site 37 /note- "potential N-glycosylation site"
FT Modified-site /note- "potential N-glycosylation site"
FT Modified-site 80 /note- "potential N-glycosylation site"
FT Modified-site /note- "potential N-glycosylation site"
FT Modified-site 108 /note- "potential N-glycosylation site"
FT Modified-site /note- "potential N-glycosylation site"
FT Modified-site 148 /note- "potential N-glycosylation site"
FT Modified-site /note- "potential N-glycosylation site"
FT Modified-site 224 /note- "potential N-glycosylation site"
FT Modified-site /note- "potential N-glycosylation site"
FT Modified-site 234 /note- "potential N-glycosylation site"
FT Modified-site /note- "potential N-glycosylation site"
FT Modified-site 235 /note- "potential N-glycosylation site"
FT Modified-site /note- "potential N-glycosylation site"
FT Modified-site 241 /note- "potential N-glycosylation site"
FT Modified-site /note- "potential N-glycosylation site"

FT Misc-difference 274 Gly residue in the formal Seq ID listing"
FT /note- "the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 353
FT /note- "the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 358
FT /note- "the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 402
FT /note- "the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 488
FT /note- "the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 496
FT /note- "the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 513
FT /note- "the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 535
FT /note- "the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 558
FT /note- "the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 567
FT /note- "the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 640
FT /note- "the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 656
FT /note- "the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 683
FT /note- "the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 685
FT /note- "the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 696
FT /note- "the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 730
FT /note- "the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 752
FT /note- "the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 763
FT /note- "the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 771
FT /note- "the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 948
FT /note- "the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 961
FT /note- "the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 982
FT /note- "the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 1032
FT /note- "the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 1089
FT /note- "the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"

FT Misc-difference 1135 /note- "the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT
Query Match
Best Local Similarity 18.8%; Score 140.5; DB 20; Length 3623;
Matches 150; Conservative 78; Mismatches 283; Indels 289; Gaps 33;
QY 70 IKSGERIVTFPCQSPENHFIIEQKIDCMGCPCEVOLOPSTS--LLPTLNFTFL- 126
DB 2306 VSSREIRYILKFTHDGSSYMGFRKXYSIASCGETVS--GDSGYIESIGYPTLPYANNVFCO 2364
QY 127 MPYKAHKSIGLEQFIPRLQIGPESCPDGT-----HSIGRIDATVVRIGTFCNGT 182
DB 2365 WFTKGLPGHYLTLSFEDFNL--QSSPG--CTKDFEITWHTSGRV-----LERTCGNST 2415
QY 183 VSRKMGQGVKALHLPFHPNRSVGFSIANSSIKRLCIIESVGE--GSATLMSAN 239
DB 2416 PSSVDYSSNV--ASYKRVTDGVSATGFRLOPKSS--RQVC-----GGDLHGPTGFTSPN 2467
QY 240 YPEGFPDELMWQFVVPYPAHLRASVSFLNPNLSN-----C 274
DB 2468 YPNPNPHARICEWTITVQEGRRIVFTNLRSTOPSCNSEHLIYENGIRSNSPLLQKLC 2527
QY 275 ERKEEVEYYIPGSTNPEVFKLEKOP-----GNMAGNFNLSGCDQDAQS 322
DB 2528 SRVNTNEKSSGNITK--VFETDGSRPYGCTASTSTEDAVCGGFLPSVSG--GNFSS 2584
QY 323 PGI-----LRLOFOVLVOHPQNESNKIYVVDLSNBRASLTIEPRPVKSRKFPVG 373
DB 2585 PGVNGIRDAYARNLDCWMTLSNPRENSISITFELISTSHQDC----- 2629
QY 374 CFVCLSEFRCSSNLITLSSSKHKISFLCDDLRLMNVNTEKTSCTDHRCCQKRSYLQYP 433
DB 2630 -FDVLEFRVGDADGPL-----IEK-----PC-----SLSAP 2654
QY 434 SDILHP-----VELHDF----- 446
DB 2655 TAPLVIPYQVWIRPVNSNERVEYTGFIIEYSPDCCGIRTDGNGVYSSPNYPLYSAMTH 2714
QY 447 -SWKLLVPRDRLSLVLPAAKIQOHTHERPCNTFSYLVASAIPSDLYGSPCGSIRK 505
DB 2715 CSMILKAPRGHTTLTSLDFLEAH----PTCSDSVTYRANGDSPSYIGRYCGSVPR 2770
QY 506 QIQVQKON--ISVLRTPAPSPQDEASRQGLTVSPFIYFKEE-----GVFTVPDTKXYVL 559
DB 2771 PIQSGSNOLIVFNT-----NNQGTQGFYATMTTNALCCGGTFHSANGT-----I 2816
QY 560 RPNMMDRGLPSLTSVSM-----NISVPRDOYACLFFFRSG----- 596
DB 2817 KSFHMPQTFPENSRCSWYITTHDSKHWETISFDSNFRIPSSDSCONSFVKWGGRLMIK 2876
QY 597 -----VVCOTGRAFMIIQOORTRAEIRP----- 620
DB 2877 TLATSCGVAPSPITYSNIFAVPQSEMAAQAQASAPISNCGTFNTSPEDITSPNF 2936
QY 621 -----LDED-----VLKPSFH-----HSHWVAINSCSPISGQ 650
DB 2937 PKQYDNNMCTYLIDADPSVLITLTVFSFLEDRAISATGTCDDHGLHI--IKGNLNSSTPL 2995
QY 651 LDLEFSVTLPRTVLDLVIL 670
DB 2996 VTIQGETLPRLVDGPAVL 3015
RESULT 13
AAB70070
ID AAB70070 standard; Protein; 595 AA.
XX
AC AAB70070;
XX
DT 14-MAY-2001 (first entry)
XX

PS Claim 1; Page 2077-2079; 2102pp; English.

XX The present invention relates to albumin fusion proteins comprising a
CC therapeutic protein X and human albumin (HA, also known as human serum
CC albumin, HSA). The proteins are useful for treating a disease or
CC disorder that may be modulated by therapeutic protein X. The albumin
CC extends the shelf-life of protein X, and may increase its biological
CC in vitro/in vivo activity. The protein is useful for treating and
CC diagnosing disorders such as cancer, reproductive disorders, digestive
CC disorders (e.g. Crohn's disease, ulcerative colitis), immune disorders
CC (e.g. acquired immunodeficiency syndrome, AIDS), endocrine disorders
CC (e.g. diabetes), haematopoietic disorders, neural disorders
CC (e.g. Alzheimer's, Parkinson's, Creutzfeldt-Jacob disease,
CC encephalomyelitis, meningitis, schizophrenia), and connective disorders
CC (e.g. osteoporosis, arthritis). ABG63326-ABG65518 represent albumin
CC fusion proteins of the invention.

XX Sequence 595 AA:

SQ Query Match 2.9%; Score 126; DB 23; Length 595;

Best Local Similarity 19.9%; Pred. No. 0.085;

Matches 135; Conservative 86; Mismatches 247; Indels 212; Gaps 34;

QY 225 SYFEEBGSKTLMSANYPPEFPEDELTMOFVYPAH---LRASVSFLNENLSCERKEERY 281
DB 33 TVQEGICVSVLCFSFSYQV-----GWTASDPVHGWFRAQ-----DHV 70
QY 282 EYVIGSTNPVFLKLEQKQPMNAGNPLSLQCGDQDQSGILRLQFOVLYQHPQNS 341
DB 71 SRNIVVATNNPARAVOE-----TRDRPH-LLGDQND 103
QY 342 NKIYVVDLSNERAMSLTEPRPVKOSRKFPVPCFV-CLESRTCSNLTLSGSKHISPL 400
DB 104 CTL-----SIRDTRESDACTVFCVGRGMKNKTKYDQLSVN-VTMS 144
QY 401 CDDLTRLMNKNVETISCTDHRVCQRRKSYSLQVPSDILH-----LPV-----ELHDF 446
DB 145 ODLSRYRLVLEPVESYTV-----OEGLCVSVCSLYLPHYNMWTASSPYGSMFEKGADI 197
QY 447 SKKLLVPDRSLVLPVPAOKLLOOHN-----EKPCNTSISYLVAASIPDOLYFGSF 498
DB 138 PVDIVPATN-----TPSGKVOEDTHGRPLLLGDPPQTNCSLSINDARKDGSKYFQ-- 249
QY 499 CPGGSIK--QIQVKNISVTLRTFAPSPFQ-----QEASRQGLTVSPFYFKEEG----- 545
DB 250 VERGSRKNMYIYDKLSVHTALTLMPTESITPGTLESQHRLNLTCS-VPMACEQGTPTTT 308
QY 546 -----VFTVTPD-TKSKVYLRTPN-WDRGLPSLT-----SVSNISVPRDQ 584
DB 309 WMGASVSSLDPTTRSSMLSLIPQPDHG-TSLTCQVTLPGAGVTMTAVRLNISVPQN 367
QY 585 VACLTFEKE-RSGVVCQGRAFMIIOEQRTAEIFSDDED----- 624
DB 368 LMTVFOGGGTSTYLRNLSALSLSLEGOSLH--VCADVSNPARLSWTGSLTSPSQS 425
QY 625 ----VLPKPSFA--HSHFWNISNCSPTSGKOLDLFSV--TLTPRYDVLVIILAAVG 675
DB 426 SMLGELPRVHVKBEGETCAQN--PLGSHISLSLSLOWEYTGKMPISGIVLGAAG 483
QY 676 GG---VLLISLALGILTCVKKKKRKTNGKPAVGIVNGNINTEMPROPKKFKGRKDNDSH 732
DB 484 GAGATVAVLFLFCIFVYVRSCKRKSAR-PAVGV--GDTGME----- 522
QY 733 VYAVIEDTMYVGHLODSSGSFLOPEVDYTRPFGCTMGVCPSPSPITCSRAPATKATEE 792
DB 523 -----DANAARV--SASGRLIESPADSPPHNAPALATSPPE--GEIYVASLSFK 572
QY 793 PPPRSPSESEPTFESHN 812
DB 573 ARPOYRQEOEALGKYESEIN 592

RESULT 15

AAE20788
ID AAE20788 standard; Protein; 3095 AA.

XX AAE20788;

AC 01-JUL-2002 (first entry)

DT Rat C3b/C4b complement receptor like protein.

DE Rat; C3b/C4b complement receptor-like molecule; immune system disorder;

KW gene therapy; rheumatoid arthritis; psoriatic arthritis; osteoarthritis;

KW inflammatory arthritis; inflammatory joint disease; Alzheimer's disease;

KW multiple sclerosis; inflammatory bowel disease; nervous system disorder;

KW transplant rejection; autoimmune disease; ischaemic condition; nocropic;

KW metabolic disorder; immunomodulatory; neuroprotective; diabetes; lupus;

KW infertility; vasodilator; obesity; cardiant.

OS Rattus rattus.

XX Key Location/Qualifiers

XX Misc-difference 196 /Label= Unknown

XX Misc-difference 847 /note= "Xaa can be any amino acid"

XX Misc-difference 847 /label= Unknown /note= "Xaa can be any amino acid"

XX WO200210199-A2.

XX 07-FEB-2002.

XX 24-JUL-2001; 2001WO-US23232.

XX 02-AUG-2000; 2000US-222504P.

XX 28-NOV-2000; 2000US-0728787.

XX (AMGE-) AMGEN INC.

XX Welcher AA, Elliott GS;

XX WPT, 2002-303934/34.

XX N-PSDB; AAD33319.

XX Nucleic acid encoding a novel C3b/C4b Complement Receptor-like nucleic

XX acid molecule, useful for treating, preventing and diagnosing

XX rheumatoid arthritis, psoriatic arthritis, inflammatory arthritis, and

XX multiple sclerosis -

XX Claim 13; Fig 3; 251pp; English.

XX The invention relates to a nucleic acid encoding a novel C3b/C4b

XX complement receptor (CR)-like nucleic acid molecule. The C3b/C4b CR-like

XX polypeptide and nucleic acid molecules may be used to treat, prevent,

XX ameliorate, diagnose and/or detect diseases such as immune system

XX disorders such as rheumatoid arthritis, psoriatic arthritis,

XX inflammatory arthritis, osteoarthritis, inflammatory joint disease,

XX autoimmune disease, multiple sclerosis, lupus, inflammatory bowel

XX disease, transplant rejection, nervous system disorders (e.g. Alzheimer's

XX disease), ischemic conditions, metabolic disorders (e.g. obesity and

XX diabetes) and infertility. The invention is useful in gene therapy. The

XX present sequence is rat C3b/C4b complement receptor like protein.

SQ Query Match 2.7%; Score 119; DB 23; Length 3095;

Best Local Similarity 17.8%; Pred. No. 4.6;

Matches 175; Conservative 89; Mismatches 296; Indels 422; Gaps 47;

QY 1 MAGLNGVSIALL---GVLLGAARLPRAEAPEI---ALPRESNTVLIKIGTPTLAK 54

DB 1225 LSGSHSGETPLPATSNOILLRPSASGASARGFHFYQAVPTSD-----T 1270

QY 55 PCYIVISKRHITMLSIKSGRIVTFPS-----COSPENHFVLE 92

Db 1271 QCSSVPEPRY-----GRIGSEFSAGSIVRECPNGYLLQGSTAIRCOQSVN-ALAQ 1321
QY 93 IQKNIDCMGCPGEGEYQLOPSTSLPTL-----NRFIWDVKAHKSIGLELOFSIPRL 146
Db 1322 WNDIIPSCVYPCS-GNFTQRRGTLISPGYEPYNNLNCWKIIVSESGIQIOVISFAT 1380
QY 147 RQIGPGSCPDGVYHISGRIDATVRIIGFCSNGTVSRIKMOEGVKMALHL----- 198
Db 1381 EQMNDSEIHGCG-----DMTAPRLGSF-SGTVPALNNTSNQCLHFQSDISVAA 1431
QY 199 PMFH-----PRN-----VSGFSTANRST----- 217
Db 1432 AGFHLEKTVGLAACQEPALPSNGIKIGDRYMNVDLSPQCEPGYTLQGRSHISCMPTV 1491
QY 218 -----KRLCI-----IESVEEGESATLMSANYPEGPEDELMTOFVPAHLRASYS 265
Db 1492 RRMVYPSPLCIATCGGTLTSM-----SGVILSPGFPGSYPNLDCIWKISLPIGYAHIQ 1546
QY 266 FLNFNLNCRKEERVE-----YY-----IPG---STINPEVFKLEDKQPGM 305
Db 1547 FLNF-----STANHDYLEIQNGPYHSPMGQFSGPDLPASLSTHETLIRFYSQHSQNR 1603
QY 306 AGNENLS-----LQGC-DQDAQSPGIL-----RLQFQ-----VLVOHPONESKTY 345
Db 1604 QG-FKLSYQAYELQNCPPAPFQNGFMINSDYSVGQISFECPGYILLGHP----- 1654
QY 346 VVDLSNERAMSLTIEPRPVKQSRKFVPGCFCLSEFRTCSSNLTLTSGSKHISF----- 399
Db 1655 -----VLTCQHTDNNMNPFRPC-----DAPCGYNTSQNGTIYSPGPEDEYPI 1699
QY 400 -----LCDDLTRLW-----MNYEKTISCTDHRXQ 424
Db 1700 LKDCMLVTPVPGHGYINFTLLQTEAVNDYIAWGDQNSPOLGVFSGNTALETAYSS 1759
QY 425 RKSYLOVPDILH---LPELHDFSWKLLVPRDLSLVLYPAKLOQHTHE----- 473
Db 1760 TNOVLLKFEHSDFSNGGFEVLFNFAFQLKRCPPPP---VPPQADLTDEDEFEIGDFVK 1814
QY 474 KPCNTSFSYL-----VASAIPSDLYFGS-----FCPGGSIK----- 505
Db 1815 YQCHPGYTLIGSDTLNCKLSQLLFQGSPTCEAQCAPANERYTESSGVILSPGPGNYFN 1874
QY 506 -----OIQVKQNISVTLTRFAPSFOE----- 527
Db 1875 SQTCAMSIKVEPNFNITL--FVDTFQSEKQFDALVFDGSSGSPLLVLSGNHTEQSNF 1932
QY 528 -----ASRQGLTVSF-IPY-----FKEGVFTVTPDT--KSKYTL----- 559
Db 1933 TSRSNHLYLKMSDHAATSKKGFKIRYAPYCSLSTLKNGVLNKTAGALGSKVQYFCKP 1992
QY 560 -----RTP-----NMDRGLPSLTSVSWNISVPRDQVACLTFEKRSGVVCQTG 602
Db 1993 GYRMIGHSNATCRNRPYGVYQWDSMAPLCAVSCGT-----PEAPGNGSFTG 2039
QY 603 RAEPIIQEQRTAEELFSLD-----EDVL-----PKPSFH-----HHS 635
Db 2040 NEFTLDSKVTYECNEGFKLDASQOATFVQEDGLMSNRGKPPCKPVPCPSIEGQLSSEHY 2099
QY 636 FMYNIS-----NCSP 645
Db 2100 LMRIVSGSINEYGAQVLLSCSP 2121

Search completed: November 11, 2002, 11:32:48
Job time : 72 secs

Db	519	ACNATCAAGGAATAANNCTGGTGTGGANGAGNCCAAAGGGGCACTTTTCCTTCAAAANTA	460
QY	5446	AGGGCTCAAAACCTCTCTGGACAAGTGTCTGGCTCC-TGACACCAGTATTTCTCGAGCTGT	5504
Db	459	AGGGCTCAAAACCTCTCTGGACAAGTGTCTGGCTCTTTAAACNCCAGTATTTCTCGAAGCTGG	400
QY	5505	GCTCTAGTGAAGGGGGCCACAGCTGAGGAACCCCTGGCTCTTTTCTTTTAAAGCCACAGCCCC	5564
Db	399	CCTCAAGTAAAGGGGGCCACACTTGAGGAACCCCTGGCTCTTTTCTTTTAAAGCCACAGCCCC	340
QY	5565	ACTTACATTAACATTTTCAGGGTCACTGCAAAACGTGAAGTGCATTTGTTGAAGCTTAC	5624
Db	339	ACTTACATTAACATTTTCAGGGTCACTGCAAAACGTGAAGTGCATTTTCTTAAAGCTTAC	280
QY	5625	TGCATGCGACGCCACTGCTCATCAGTGGTGTGGCCATGCTACGAGGAAGGCCAGCGCA	5684
Db	279	TGCATGCGACGCCACTGCTCATCAGTGGTGTGGCCATGCTTACGAGGAAGGCCAGCGCA	220
QY	5685	TGCAGACGTGTCTTAAATGCTGTGGTCAATTGCACAGAAAGGAAAGGTCTCAAGGAAGAG	5744
Db	219	TGCAGGACTGGTCTCTAAATGCTGTGGTCAATTGCACAGAAAGGAAAGGTCTCAAGGAAGAG	160
QY	5745	TCAACTGGGCAACGCAACAACCCACCGGACATGGCTGTAAAGTTAGCAGACTGGTG	5804
Db	159	TCAACTGGGCAACGCAACAACCCACCGGACATGGCTGTGTAAAGTTAGCAGACTGGTG	100
QY	5805	TGTGTGATCTGCGAGTGCCTTCACTGGAATAATATTATTCATTGCAAGATCTTTTAGGTG	5864
Db	99	TGTGTGATCTGCGAGTGCCTTCACTGGAATAATATTATTCATTGCAAGATCTTTTAGGTG	40
QY	5865	GCATTTATTCATTTCTCTGTGCTTTAAATTAACAAATGT	5903
Db	39	GCATTTATTCATTTCTCTGTGCTTTAAATTAACAAATGT	1

```

RESULT 2
US-09-798-096-10/c
: Sequence 10, Application US/09798096
: Patent No. 639378
: GENERAL INFORMATION:
: APPLICANT: Donna T. Ward
: APPLICANT: Andrew T. Walt
: TITLE OF INVENTION: ANTISENSE MODULATION OF RECOL2 EXPRESSION
: FILE REFERENCE: RTS-0207
: CURRENT APPLICATION NUMBER: US/09/798,096
: CURRENT FILING DATE: 2001-03-01
: NUMBER OF SEQ ID NOS: 89
: SEQ ID NO 10
: LENGTH: 99500
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: US-09-798-096-10

```

Query Match	3.5%	Score 218.6;	DB 4;	Length 99500;
Best Local Similarity	82.6%;	Pred. No. 1.3e-41;		
Matches 262;	Conservative	0;	Mismatches 54;	Indels 1;
			Gaps	1

[illegible]

OY	4731	GGGCCCAAGCATCTCCTGCGTGGCCGCCCAAAGTGCTGGATTACGAATAGTAGCCA	4790
Dδ	65942	GACCACAAGCATTCACACTTTGCCTTGTGCTCCCAGAAGTGTGGATTACGACATAGCCA	65883
OY	4791	CCGCATCCAGCCCCA	4807
Dδ	65882	CCGGGCCCAAGCCTAACCA	65866

```

: RESULT 3
: US-08-646-301A-1/C
: Sequence 1, Application US/08646301A
: Patent No. 6194211
: GENERAL INFORMATION:
: APPLICANT: Richards, Cynthia Ann
: APPLICANT: Huber, Brian E.
: TITLE OF INVENTION: Transcriptional Regulatory Sequence of Carcinoembryonic
: Patent No. 6194211
: TITLE OF INVENTION: Antigen for Expression Targeting
: FILE REFERENCE: PBI50805W
: CURRENT APPLICATION NUMBER: US/08/646,301A
: CURRENT FILING DATE: 1996-05-16
: NUMBER OF SEQ ID NOS: 25
: SOFTWARE: Patentln Ver. 2.1
: SEQ ID NO 1
: LENGTH: 11288
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-08-646-301A-1

```

Query Match	3.5%	Score 218.2;	DB 4;	Length 11288;
Best Local Similarity	81.7%;	Pred. No. 5.8e-42;		
Matches 264; Conservative	0;	Mismatches 58;	Indels 1;	Gaps 1

QY	4480	TTCCTTTTTTCTTTCTTTCTTTCTTTTTTTTTTAATGACACAGCAATCATCT	4539
Dd	858	TTTTTTTTTTCTTTCTTTCTTTTTTTTTTTTTTTTAAAGACTGGCCT	799
QY	4540	GTTCCTAGCGCTGAGNGCAGTGCGGCATCTCGGTCACTGCACACTCGCTCGG	4599
Dd	798	GTTBGCCAAGCTSGAGTGCAGTGGCAAATTCTTGGCTACTGCACAACCTCTGCTTCGG	739
QY	4600	CTCAAGCAATTCCTCCACCTGCACGCTCCCAATAATAGCTGGATCTGGCACAAAACCA	4659
Dd	738	TCCAGCAATTTGTCACGCTCACCCTCCGAGTAGCTGGGATTAAGGCGGCGCACACCA	679
QY	4660	TGCCCCAGCTAA-TTTGTATTTTTTTGTAGAAGACAGGGTTCCACATGTTGCCACGCTGG	4718
Dd	678	TGCCACACTAATTTTTTGTATTTTATAGACCGGGGTTCCACATATTTGGCCAGGGTG	619
QY	4719	TCTCAACTCTTGGGCTCAAGCAATCTCTCTGCTCGGCTTCCCAAATGCTGGGATTAC	4778
Dd	618	TCTTGAACCTCTCAACCTCAAGTAACTGTGCCACCTCAGCCTTCTAAAGTGTGGGATTAC	559
QY	4779	AGATGTAGCAGCCGATCAGC-4801	
Dd	558	AGGGTGAGCCACCGCGCCGGG-536	

```

RESULT 4
US-08-481-968A-4/C
: Sequence 4, Application US/08481968A
: Patent No. 6300490
: GENERAL INFORMATION
:
: APPLICANT: Huber, Brian
: APPLICANT: Richards, Cynthia
: TITLE OF INVENTION: Molecular Constructs Comprising a Carcinoembryonic Antigen (
: FILE REFERENCE: PB1087054
: CURRENT APPLICATION NUMBER: US/08/481,968A
: CURRENT FILING DATE: 1998-06-07
: NUMBER OF SEQ ID NOS: 36

```

```

; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 11288
; TYPE: DNA
; ORGANISM: Homo sapiens
US-08-481-968A-4

```

Query Match	3.5%	Score 218.2;	DB 4;	Length 11288;
Best Local Similarity	81.7%;	Pred. NO. 5.8e-42;		
Matches 264; Conservative	0;	Mismatches 58;	Indels 1;	Gaps 1;

[illegible]

```

RESULT 5
US-08-154-712B-4/C
: Sequence 4, Application US/08154712B
: Patent No. 6337209
: GENERAL INFORMATION:
: APPLICANT: Huber, Brian
: APPLICANT: Richards, Cynthia
: TITLE OF INVENTION: Molecular Constructs Containing a Carcinoembryonic Antigen Reg
: TITLE OF INVENTION: Sequence
: FILE REFERENCE: PB10870S3
: CURRENT APPLICATION NUMBER: US/08/154,712B
: CURRENT FILING DATE: 1993-11-19
: NUMBER OF SEQ ID NOS: 36
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 4
: LENGTH: 11288
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-08-154-712B-4

```

	Query Match	3.5%	Score 218.2	DB 4:	Length 11288;
	Best Local Similarity	81.7%	Pred. No. 5.8e-42;		
	Matches 264;	Conservative 0;	Mismatches 36;	Indels 1;	Gaps 1;
QY	4480	TTTTCTTTTTTTTTCTTTTTCTTTTTCTTTTTTAAATGAGACAGAGATCATCT	4539		
Db	858	TTTTCTTTTTCTTTCTTTCTTTTTTTTTTTTTTTTTTTTTTAAAGAGTGGGCT	799		
QY	4540	GTTGCTTAGGCTGAGAGTGCATGAGGGCAATCGGGCTACATGCAACCTCTGGCTCTGGG	4599		
Db	798	GTTGGCCCAAGTGGAGTGCAGATGGGACAACTCTGGCTACATGCAACCTCTGGCTCTGGG	739		
QY	4600	CTCAAGCAATTCCTCCACCTCAGCCCTCCCAAAATAGCTGGAGTACGTGGCACAACACCA	4659		
Db	738	TCCAAAGCAATGTTCAGGCTCAGGCTCCCGAGTACTGGGATTAAGAGCGGCACACCA	679		
QY	4660	TGCCAGCTAA-TTTTTGATTTTTTTTGTAGAGACAGGTTTCCACATGTTGCCAGGCTGG	4718		

D _b	678	TGCCACGATATTTTTGTATTTTAGTAGAGCCGGGGTTTCACCATATTGGCCAGGGTGG	619
Q _y	4719	TCTCAACCTCCTGGGGCTCAAGCATCCTCCTGCCTGGCGCTCCCAAGTGCTGGGATTAC	4778
D _b	618	TCTTGAACCTCCTGACCTCAAGTATATGTGCCACCCTCAGCGTTCTTAAGTGCTGGGATTAC	559
Q _y	4779	AGATGTGAGCCACCGCATCCAGC	4801
D _b	558	AGGCGTGAGCCACCGCCCGGG	536

```

: RESULT 6
: US-09-167-681-45
: Sequence 45, Application US/09167681A
: Patent No. 6265561
: GENERAL INFORMATION:
: APPLICANT: Weinsteinbaum, M.D., Richard M.
: APPLICANT: Ratfogianis, Rebecca B.
: APPLICANT: Wood, Thomas C.
: APPLICANT: O'leanness, Diane M.
: TITLE OF INVENTION: SULFOTRANSFERASE SEQUENCE VARIANTS
: FILE REFERENCE: 07039/118001
: CURRENT APPLICATION NUMBER: US/09/167,681A
: CURRENT FILING DATE: 1998-10-07
: NUMBER OF SEQ ID NOS: 52
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 45

```

```

? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (4361)..(4507)
? NAME/KEY: CDS
? LOCATION: (4612)..(4737)
? NAME/KEY: CDS
? LOCATION: (4827)..(4925)
? NAME/KEY: CDS
? LOCATION: (6332)..(6447)
? NAME/KEY: CDS
? LOCATION: (6543)..(6638)
? NAME/KEY: CDS
? LOCATION: (7137)..(7316)
? NAME/KEY: CDS
? LOCATION: (7439)..(7553)
? JS-09-167-681-45

```

[illegible]

PRIOR FILING DATE: 1992-06-04
PRIOR APPLICATION NUMBER: 07/952,796
PRIOR FILING DATE: 1992-09-24
NUMBER OF SEQ ID NOS: 34
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 5
LENGTH: 5262
TYPE: DNA
ORGANISM: HUMAN
FEATURE:
OTHER INFORMATION: EXON 35-161; EXON 1142-1297; EXON 1984-2187;
OTHER INFORMATION: EXON 5170-5255; INTRON 162-1141; INTRON
OTHER INFORMATION: 1298-1983; INTRON 2188-5169; CDS 35-161; CDS
OTHER INFORMATION: 1142-1297; CDS 1984-2187; CDS 5170-5255
NAME/KEY: exon
LOCATION: (35)..(160)
NAME/KEY: exon
LOCATION: (1142)..(1297)
NAME/KEY: exon
LOCATION: (1984)..(2187)
NAME/KEY: exon
LOCATION: (5170)..(5256)
NAME/KEY: Intron
LOCATION: (162)..(1141)
NAME/KEY: Intron
LOCATION: (1298)..(1983)
NAME/KEY: Intron
LOCATION: (2188)..(5169)
OTHER INFORMATION: n = a or g or t or c, any base
US-08-520-373D-5

Query Match 3.5%; Score 216.8; DB 4; Length 5262;
Best Local Similarity 82.8%; Pred. No. 8.6e-42;
Matches 256; Conservative 0; Mismatches 52; Indels 1; Gaps 1;

QY 4495 TTTTCTTTCTTTCTTTTATGAGACAGATCTCTCTGCTAGGCTGGA 4554
DB 564 TTTTCTTTCTTTCTTTTNNCTTCTGAGACGAGTCTGCTTGTGNCNCAAGCTGGA 623
QY 4555 GTGACGTGGGCAATCTGCGCTCACTGCAACCTCTGCTGCGGCTCAAGCAATTTCTCC 4614
DB 624 GTGACGTGGGATGATCTGAGTCTGCAACCTCTGCTGCGGCTCAAGCAATTTCTCC 683
QY 4615 CACCTCAGCTCCCAATAGCTGGGATCTGCGACAAACCACTCCCAAGCTAA-TTT 4673
DB 684 TGCTCAGCTCCCAAGATAGCTGGGATCTGCGACCACTCCCAAGCTAAATTTT 743
QY 4674 TGTATTTTGTAGAGACAGGTTTCAACATGTTGCCAGGCTGCTCAACCTCCGCGG 4733
DB 744 TGTATTTTGTAGAGACAGGTTTCAACATGTTGCCAGGCTGCTCAACCTCCGCGG 803
QY 4734 CTCAAGCAATCTCTGCTGCGCTCCCAAGTGTGGGATTACAGATGTGAGCCACCG 4793
DB 804 CTCAAGGATCCACCCGCTGCGCTCCCAAGTGTGGGATTACAGGCGTGAACCCACCG 863
QY 4794 CATCCAGCC 4802
DB 864 CGCCTGGCC 872

RESULT 15
US-08-367-841A-43
Sequence 43, Application US/08367841A
Patent No. 6319687
GENERAL INFORMATION:
APPLICANT: Chader, Gerald J.; Rodriguez,
APPLICANT: Ignacio R.; Mazuruk, Krzysztof;
APPLICANT: Tomdrian-Tink, Joyce
TITLE OF INVENTION: PIGMENT EPITHELIAL
TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION GENOMIC
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:

ADDRESSEE: Morgan & Flinnegan
STREET: 345 Park Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/367,841A
FILING DATE: 30-DEC-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/257,963
FILING DATE: 07-JUN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/952,796
FILING DATE: 24-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: DOROTHY R. AUTH
REGISTRATION NUMBER: 36434
REFERENCE/DOCKET NUMBER: 20264126052
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 22481 Base Pairs
TYPE: Nucleic Acid
STRANDEDNESS: Double
TOPOLOGY: Unknown
MOLECULE TYPE: Genomic DNA
FEATURE:
NAME/KEY: P1-147
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION: full length genomic
OTHER INFORMATION: sequence for PEDF plus flanking sequences.
US-08-367-841A-43

Query Match 3.5%; Score 216.8; DB 4; Length 22481;
Best Local Similarity 82.8%; Pred. No. 1.7e-41;
Matches 256; Conservative 0; Mismatches 52; Indels 1; Gaps 1;

QY 4495 TTTTCTTTCTTTCTTTTATGAGACAGATCTCTCTGCTAGGCTGGA 4554
DB 15129 TTTTCTTTCTTTCTTTTNNCTTCTGAGACGAGTCTGCTTGTGNCNCAAGCTGGA 15188
QY 4555 GTGACGTGGGCAATCTGCGCTCACTGCAACCTCTGCTGCGGCTCAAGCAATTTCTCC 4614
DB 15189 GTGACGTGGGATGATCTGAGTCTGCAACCTCTGCTGCGGCTCAAGCAATTTCTCC 15248
QY 4615 CACCTCAGCTCCCAATAGCTGGGATCTGCGACAAACCACTCCCAAGCTAA-TTT 4673
DB 15249 TGCTCAGCTCCCAAGATAGCTGGGATCTGCGACCACTCCCAAGCTAAATTTT 15308
QY 4674 TGTATTTTGTAGAGACAGGTTTCAACATGTTGCCAGGCTGCTCAACCTCCGCGG 4733
DB 15309 TGTATTTTGTAGAGACAGGTTTCAACATGTTGCCAGGCTGCTCAACCTCCGCGG 15368
QY 4734 CTCAAGCAATCTCTGCTGCGCTCCCAAGTGTGGGATTACAGATGTGAGCCACCG 4793
DB 15369 CTCAAGGATCCACCCGCTGCGCTCCCAAGTGTGGGATTACAGGCGTGAACCCACCG 15428
QY 4794 CATCCAGCC 4802
DB 15429 CGCCTGGCC 15437

Search completed: November 11, 2002, 11:36:24

Tue Nov 12 18:03:24 2002

Job time : 724 secs

us-09-899-569a-3.rni

This Page Blank (uspto)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 11, 2002, 04:57:15 ; Search time 190 Seconds
(without alignments)
11508.120 Million cell updates/sec

Title: US-09-899-569A-3

Perfect score: 6163
Sequence: 1 ccacgcgcgcattgggagt.....agtcgcgcaaaaaaaaaa 6163

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 320260 seqs, 177392727 residues

Total number of hits satisfying chosen parameters: 640520

```
Minimum DB seq length: 0
```

Maximum DB seq length: 200000000000

Post-processing: Minimum Match 08

Listing first 45 summaries

Listing first 45 summaries

```

1: /cgn2_6/prodatae1/pubnpa/us07_PUBCOMB.seq.*
2: /cgn2_6/prodatae1/pubnpa/PCT_NEW_PUB.seq.*
3: /cgn2_6/prodatae1/pubnpa/us06_NEW_PUB.seq.*
4: /cgn2_6/prodatae1/pubnpa/us06_PUBCOMB.seq.*
5: /cgn2_6/prodatae1/pubnpa/us07_NEW_PUB.seq.*
6: /cgn2_6/prodatae1/pubnpa/PCUS_PUBCOMB.seq.*
7: /cgn2_6/prodatae1/pubnpa/us08_NEW_PUB.seq.*
8: /cgn2_6/prodatae1/pubnpa/us08_PUBCOMB.seq.*
9: /cgn2_6/prodatae1/pubnpa/us09_NEW_PUB.seq.*
10: /cgn2_6/prodatae1/pubnpa/us09_PUBCOMB.seq.*
11: /cgn2_6/prodatae1/pubnpa/us10_NEW_PUB.seq.*
12: /cgn2_6/prodatae1/pubnpa/us10_PUBCOMB.seq.*
13: /cgn2_6/prodatae1/pubnpa/us60_NEW_PUB.seq.*
14: /cgn2_6/prodatae1/pubnpa/us60_PUBCOMB.seq.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	61.63	100.0	6163	10	US-09-899-569A-3	Sequence 3, April
2	5854.2	95.0	5897	10	US-09-899-569A-1	Sequence 1, April
3	5442.4	8.8	544	10	US-09-923-217-362	Sequence 362, App
4	5442.4	8.8	544	10	US-09-833-263-362	Sequence 362, App
5	533	8.6	541	10	US-09-813-343-135	Sequence 135, App
6	522.6	8.5	636	10	US-09-813-343-666	Sequence 666, App
7	489.2	7.9	636	10	US-09-879-536-370	Sequence 370, App
8	482	7.8	535	10	US-09-815-343-393	Sequence 393, App
9	347	5.6	348	10	US-09-867-701-3051	Sequence 3051, App
10	339	5.5	339	10	US-09-867-701-9866	Sequence 9866, App
11	331	3.7	5157	10	US-09-764-877-2605	Sequence 2605, App
12	231	3.7	5199	10	US-09-764-877-2607	Sequence 2607, App
13	230.8	3.7	167343	10	US-09-962-436-281	Sequence 281, App
14	230.8	3.7	167343	10	US-09-964-824A-273	Sequence 273, App
15	226	3.7	58637	10	US-09-982-091A-5	Sequence 5, April
16	223.6	3.6	7662	10	US-09-764-860-1109	Sequence 1109, App
17	222.6	3.6	32170	10	US-09-764-860-1108	Sequence 1108, App
18	222.4	3.6	1400	10	US-09-263-959-235	Sequence 295, App
19	222.4	3.6	172637	10	US-09-805-458A-3	Sequence 3, April

C 20	222	3	6	68493	10	US-09-265-99-1	Sequence 1, Appl1
C 21	221.8	3	6	98865	10	US-09-770-609A-3	Sequence 3, Appl1
C 22	221	3	6	20029	10	US-09-764-84-7-1624	Sequence 1624, Appl1
C 23	220.8	3	6	5564	10	US-09-764-84-7-1207	Sequence 1207, Appl1
C 24	220.8	3	6	12149	10	US-09-764-86-9-2258	Sequence 2258, Appl1
C 25	220.8	3	6	30515	10	US-09-764-84-7-1208	Sequence 1208, Appl1
C 26	220.6	3	6	17397	10	US-09-764-86-9-1945	Sequence 1945, Appl1
C 27	220.6	3	6	19334	10	US-09-764-86-9-1943	Sequence 1943, Appl1
C 28	220.6	3	6	19345	10	US-09-764-86-9-1944	Sequence 1944, Appl1
C 29	219.6	3	6	26048	10	US-09-764-86-9-1556	Sequence 1556, Appl1
C 30	219.4	3	6	262	10	US-09-815-313-903	Sequence 903, Appl1
C 31	219.4	3	6	9745	10	US-09-764-86-9-2259	Sequence 2259, Appl1
C 32	219.4	3	6	44848	10	US-09-776-874A-42	Sequence 42, Appl1
C 33	219.4	3	6	45845	10	US-09-927-091-6	Sequence 6, Appl1
C 34	219.4	3	6	465237	10	US-09-933-267A-1	Sequence 1, Appl1
C 35	219	3	6	10953	10	US-09-764-84-6-313	Sequence 313, Appl1
C 36	218.4	3	5	57130	10	US-09-833-081-3	Sequence 3, Appl1
C 37	218.2	3	5	11288	10	US-09-947-925A-4	Sequence 4, Appl1
C 38	218.2	3	5	11288	10	US-09-822-634-1	Sequence 1, Appl1
C 39	218.2	3	5	15056	10	US-09-397-892-4	Sequence 4, Appl1
C 40	217.6	3	5	8447	10	US-09-934-426-543	Sequence 543, Appl1
C 41	217.6	3	5	8447	10	US-09-880-107-3320	Sequence 3320, Appl1
C 42	217.4	3	5	31594	10	US-09-764-86-0-599	Sequence 599, Appl1
C 43	217.4	3	5	30175	10	US-09-78-878-3	Sequence 3, Appl1
C 44	217.4	3	5	30175	12	US-10-163-381-3	Sequence 3, Appl1
C 45	217.2	3	5	32199	10	US-09-764-85-5-110	Sequence 210, Appl1

ALIGNMENTS

```

RESULT 1
US-09-899-569A-3
Sequence 3, Application US/09899569A
Patent No. US20020142003A1
GENERAL INFORMATION:
APPLICANT: No. US20020142003Albert Schweifer
APPLICANT: Marwa Scheel-Mostagier
APPLICANT: Wolfgang Sommergruber
APPLICANT: Roger Abseher
TITLE OF INVENTION: Tumorsozialisiertes Antigen (B345)
FILE REFERENCE: 0657.2280001
CURRENT APPLICATION NUMBER: US/09/899,569A
CURRENT FILING DATE: 2001-07-06
PRIOR APPLICATION NUMBER: DE 100 33 080.0
PRIOR FILING DATE: 2000-07-07
PRIOR APPLICATION NUMBER: DE 101 19 294.0
PRIOR FILING DATE: 2001-04-19
PRIOR APPLICATION NUMBER: US 60/243,158
PRIOR FILING DATE: 2000-10-25
PRIOR APPLICATION NUMBER: US 60/297,747
PRIOR FILING DATE: 2001-06-14
NUMBER OF SEQ ID NOS: 40
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3
LENGTH: 6163
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: 5' UTR
LOCATION: (1)..(282)
NAME/KEY: GC signal
LOCATION: (147)..(157)
NAME/KEY: misc.feature
LOCATION: (201)..(209)
OTHER INFORMATION: cap signal; Transcription start
NAME/KEY: 3' UTR
LOCATION: (2794)..(6163)
NAME/KEY: 3' UTR
LOCATION: (2794)..(6163)
NAME/KEY: CDS
LOCATION: (283)..(2793)
US-09-899-569A-3

```


|||||
Db 2101 ATCATCAGAGACGAGCGGAGCCGGGCTGAGGAGATCTTCAGCCTGAGACGAGATGTCCTC 2160
OY 2161 CCCAAGCCAAAGCTTCACCATCATCAGAGCTTCGGGTCACATCTCTACTGCGAGCCCCAG 2220
|||||
Db 2161 CCCAAGCCAAAGCTTCACCATCATCAGAGCTTCGGGTCACATCTCTACTGCGAGCCCCAG 2220
OY 2221 ACGGCAACGACACTAGACCTGGCTCTTCGCGTGACACTTACCCCAAGGAGCTGGAGCTTG 2280
|||||
Db 2221 ACGGCAACGACACTAGACCTGGCTCTTCGCGTGACACTTACCCCAAGGAGCTGGAGCTTG 2280
OY 2281 ACTGTCATCTCATGCGACGCGTGAGGAGTGAGTCTTACTGCTGTCTGCCCCGAGGCTC 2340
|||||
Db 2281 ACTGTCATCTCATGCGACGCGTGAGGAGTGAGTCTTACTGCTGTCTGCCCCGAGGCTC 2340
OY 2341 ATCATTTGCTGTGTGAAAAAGAAAGAAAGAACAAACAGAGGCCCGCTGTGGGTATC 2400
|||||
Db 2341 ATCATTTGCTGTGTGAAAAAGAAAGAAAGAACAAACAGAGGCCCGCTGTGGGTATC 2400
OY 2401 TACAAATGGCAATCAATATCTGATGCGGAGGACGCCAAAAAGTTTGAGAAAGGGCGA 2460
|||||
Db 2401 TACAAATGGCAATCAATATCTGATGCGGAGGACGCCAAAAAGTTTGAGAAAGGGCGA 2460
OY 2461 AAGGACAATGACTCCCATGTGTATGCACTCATGAGACACATGCTATATGCGGACATCTG 2520
|||||
Db 2461 AAGGACAATGACTCCCATGTGTATGCACTCATGAGACACATGCTATATGCGGACATCTG 2520
OY 2521 CTACAGAGATTCCAGCGGCTCTCTTCGACGCCAGAGGTGAGACCTTACCGCGCTTCCAG 2580
|||||
Db 2521 CTACAGAGATTCCAGCGGCTCTCTTCGACGCCAGAGGTGAGACCTTACCGCGCTTCCAG 2580
OY 2581 GGCACCATGGGGGTCTGTCCTCCCTCCACCCACCATATGTCGCGAGGCCCCCAACTGCA 2640
|||||
Db 2581 GGCACCATGGGGGTCTGTCCTCCCTCCACCCACCATATGTCGCGAGGCCCCCAACTGCA 2640
OY 2641 AAGTTGGCCACTGAGAGGACCACTCTCGCTCCCTCTCTGAGTCTGAGATGTAACCGTAC 2700
|||||
Db 2641 AAGTTGGCCACTGAGAGGACCACTCTCGCTCCCTCTCTGAGTCTGAGATGTAACCGTAC 2700
OY 2701 ACCTTCTCCCATCCCAACAAATGGGAGTGTAGCAGCAGACACAGACATTTCCCTTACTG 2760
|||||
Db 2701 ACCTTCTCCCATCCCAACAAATGGGAGTGTAGCAGCAGACACAGACATTTCCCTTACTG 2760
OY 2761 AACACTCAGAGGCCCATGAGCGACAGAGATAACTTGATCCACTTCAGACGCTTTCCTGA 2820
|||||
Db 2761 AACACTCAGAGGCCCATGAGCGACAGAGATAACTTGATCCACTTCAGACGCTTTCCTGA 2820
OY 2821 GTTTCATTAAGCAGGCGCACTGAGACACCGCTCGTGTCTTAAACGAAATCTTAAGAA 2880
|||||
Db 2821 GTTTCATTAAGCAGGCGCACTGAGACACCGCTCGTGTCTTAAACGAAATCTTAAGAA 2880
OY 2881 GAGGAATTAATACAGAAAGAAACAGCAGAGAGTTTCTCTGACACCGCCAACTTCACATTGC 2940
|||||
Db 2881 GAGGAATTAATACAGAAAGAAACAGCAGAGAGTTTCTCTGACACCGCCAACTTCACATTGC 2940
OY 2941 TTAGTGGAGTCAATTCCTTAAGGGCAAGACATTTGAATAATGATGAATTCGAATCTGGATCAGT 3000
|||||
Db 2941 TTAGTGGAGTCAATTCCTTAAGGGCAAGACATTTGAATAATGATGAATTCGAATCTGGATCAGT 3000
OY 3001 CATGACAGCTCATGTGCTCTCAACTTAAGCTGTGCGGTGTTAGGACAGCTGTAAATGAGAG 3060
|||||
Db 3001 CATGACAGCTCATGTGCTCTCAACTTAAGCTGTGCGGTGTTAGGACAGCTGTAAATGAGAG 3060
OY 3061 AGAGAGGCTGATGATCACTAGACATAGGAGTTGACAGCAAGCCCTGAGATTGAGAGTGTAAAC 3120
|||||
Db 3061 AGAGAGGCTGATGATCACTAGACATAGGAGTTGACAGCAAGCCCTGAGATTGAGAGTGTAAAC 3120
OY 3121 AAGAGCTTGCCCTTTACAGAGACAAAGTTTCCAATTCCAAGGAGCTTACTGAGGTCCCTA 3180
|||||
Db 3121 AAGAGCTTGCCCTTTACAGAGACAAAGTTTCCAATTCCAAGGAGCTTACTGAGGTCCCTA 3180
OY 3181 CTCTACTGGGGTCCCAAGATGAAAAAGCAATGTCCTTTTATATTTATTTATTTG 3240
|||||

|||||
Db 3181 CTCTACTGGGGTCCCAAGATGAAAAAGCAATGTGCCCTTTTATTTATTTATTTG 3240
OY 3241 TGGTCTGTGTATTTTAAGAGATCAAAATGATATACCACCTAGCTCTTTTACTGACTTA 3300
|||||
Db 3241 TGGTCTGTGTATTTTAAGAGATCAAAATGATATACCACCTAGCTCTTTTACTGACTTA 3300
OY 3301 GTAATTAACATATCTTAACCTGTTTGGATGCTGGGTTGTGACCTTTCACAGCCGCTAGAT 3360
|||||
Db 3301 GTAATTAACATATCTTAACCTGTTTGGATGCTGGGTTGTGACCTTTCACAGCCGCTAGAT 3360
OY 3361 AAACGTGTGCTGTCCCGCAGGTGTGGGAATTAATTAACATCTGTCCAACGAAAGA 3420
|||||
Db 3361 AAACGTGTGCTGTCCCGCAGGTGTGGGAATTAATTAACATCTGTCCAACGAAAGA 3420
OY 3421 ATGTGTGTGTTTACAGACATTTACACATATCTGCTTTGATTAAGAGACTTCTGATTTCTC 3480
|||||
Db 3421 ATGTGTGTGTTTACAGACATTTACACATATCTGCTTTGATTAAGAGACTTCTGATTTCTC 3480
OY 3481 TAGTGGGTGCGGTGTTATCCCATTTGTGGAATTCATCTTGATGCCATTGTCTTATAGT 3540
|||||
Db 3481 TAGTGGGTGCGGTGTTATCCCATTTGTGGAATTCATCTTGATGCCATTGTCTTATAGT 3540
OY 3541 CCTAGCAATTAAGAGAAATTTCTCAAGTTTCCATGTGCGGTTCTCTAGTCGAGCAATA 3600
|||||
Db 3541 CCTAGCAATTAAGAGAAATTTCTCAAGTTTCCATGTGCGGTTCTCTAGTCGAGCAATA 3600
OY 3601 CTTTGACATTTAAAGAGAAATTTAGAGAAATATCTCATCTCTTAAAAATGTTTAAATATA 3660
|||||
Db 3601 CTTTGACATTTAAAGAGAAATTTAGAGAAATATCTCATCTCTTAAAAATGTTTAAATATA 3660
OY 3661 TACCAACAGTGGCCCCCTGACATTAATTTCTGTCGCACTGCAACCAATTAATGTTG 3720
|||||
Db 3661 TACCAACAGTGGCCCCCTGACATTAATTTCTGTCGCACTGCAACCAATTAATGTTG 3720
OY 3721 CTTTAAACCAACATTAAGCTTATAGTCTCGGGGATCAGAAATTCAAAAATGATGTCCCT 3780
|||||
Db 3721 CTTTAAACCAACATTAAGCTTATAGTCTCGGGGATCAGAAATTCAAAAATGATGTCCCT 3780
OY 3781 GAATGAATAATCAAGGTGTACAGAGAGCTGTCTCTCTGAGGCTCTAGGAGAAAGCG 3840
|||||
Db 3781 GAATGAATAATCAAGGTGTGTACAGAGAGCTGTCTCTCTGAGGCTCTAGGAGAAAGCG 3840
OY 3841 GTTCCTTGCAATTTCAAGGCTTCTAGAGGCTGTGATGCCATGCCAGGCTCCAGTGGCTGTC 3900
|||||
Db 3841 GTTCCTTGCAATTTCAAGGCTTCTAGAGGCTGTGATGCCATGCCAGGCTCCAGTGGCTGTC 3900
OY 3901 AAGCTTTTCTCAGATGGCATCACTGTGACACTGTGCCCTCCCACTTCCCTTTGACTTAC 3960
|||||
Db 3901 AAGCTTTTCTCAGATGGCATCACTGTGACACTGTGCCCTCCCACTTCCCTTTGACTTAC 3960
OY 3961 AAAGCCACAGAGAAATCCAGAGATATCTCTCACTTAAAGATCCCTTCATCTCCTGGA 4020
|||||
Db 3961 AAAGCCACAGAGAAATCCAGAGATATCTCTCACTTAAAGATCCCTTCATCTCCTGGA 4020
OY 4021 AGAGCTTTTGGCCATGCAAGACAAACATATGCCACAGAGGTGGGATTAAGACACAGACATCTT 4080
|||||
Db 4021 AGAGCTTTTGGCCATGCAAGACAAACATATGCCACAGAGGTGGGATTAAGACACAGACATCTT 4080
OY 4081 TGGGGTGTGTATTTATGCTTACACACCTTCTGTGCACATGACTTCCACAGAGAGGCTA 4140
|||||
Db 4081 TGGGGTGTGTATTTATGCTTACACACCTTCTGTGCACATGACTTCCACAGAGAGGCTA 4140
OY 4141 CAAATATGATCTGGCGACAGGAGTGTGTTAGCTTGCGGACTGTAACTTAAAAA 4200
|||||
Db 4141 CAAATATGATCTGGCGACAGGAGTGTGTTAGCTTGCGGACTGTAACTTAAAAA 4200
OY 4201 ACCCGATTCAGAAATCTGGCCATGCTGGGCTGTACATTTCTACCTAGCAACAATGGGC 4260
|||||
Db 4201 ACCCGATTCAGAAATCTGGCCATGCTGGGCTGTACATTTCTACCTAGCAACAATGGGC 4260
OY 4261 TGGAGCTGGGCAACCACTGTGCCCTTGAAGAGGGGTGTCCACTCACCAGGTCCACAGC 4320
|||||
Db 4261 TGGAGCTGGGCAACCACTGTGCCCTTGAAGAGGGGTGTCCACTTCCACAGGTCCACAGC 4320

OY	4321	CCACACATACGCCCTATACCTTCCACACATGAGGCTAAGTGTGTTCTTACTGATCAATG	4380
Db	4321	CCACACATCGGCCCTATACCTTCCACACATGAGGCTAAGTGTGTTCTTACTGATCAATG	4380
OY	4381	CCCCGACGGTTGCAATTTATGTAATGAAGAAAGAAAGACGTGGGATTAATCCTATACG	4440
Db	4381	CCCCGACGGTTGCAATTTATGTAATGAAGAAAGAAAGACGTGGGATTAATCCTATACG	4440
OY	4441	TGAGTAGACATGAGACCAATGTGTGCTACATATACCCTTTTCTTTTCTTTTCTTCT	4500
Db	4441	TGAGTAGACATGAGACCAATGTGTGCTACATATACCCTTTTCTTTTCTTTTCTTCT	4500
OY	4501	TTTTCTTTTTTTTTTTAATGTGACACAGACTCTCAATCTGTGTGCTAGGCTGAGTACG	4560
Db	4501	TTTTCTTTTTTTTTTTTTAATGTGACACAGACTCTCAATCTGTGTGCTAGGCTGAGTACG	4560
OY	4561	TGGGGCAATCTGGCTCTACGACAAACGCTCCGCGGGCTCAACGATATCTCCACCTC	4620
Db	4561	TGGGGCAATCTGGCTCTACGACAAACGCTCCGCGGGCTCAACGATATCTCCACCTC	4620
OY	4621	AGCCCTCCAAATAGCTGGATCACTGGCAAAACCAATGCGCAGCTAATTTGTATTT	4680
Db	4621	AGCCCTCCAAATAGCTGGATCACTGGCAAAACCAATGCGCAGCTAATTTGTATTT	4680
OY	4681	TTTTGTAGAGACAGGGTTTCAACATGTGTGCGCAGGCTGTCAACCTCTGGGCTCAAGC	4740
Db	4681	TTTTGTAGAGACAGGGTTTCAACATGTGTGCGCAGGCTGTCAACCTCTGGGCTCAAGC	4740
OY	4741	AATCCTCTGCTGCGCTCCCAAAATGCTGGGATTAACAGATGTAGACACACGATCCAG	4800
Db	4741	AATCCTCTGCTGCGCTCCCAAAATGCTGGGATTAACAGATGTAGACACACGATCCAG	4800
OY	4801	CCCCACACCTCATTTATACCAATTACCTGCCAGTACTGTGACCTTTGCTTCTCAC	4860
Db	4801	CCCCACACCTCATTTATACCAATTACCTGCCAGTACTGTGACCTTTGCTTCTCAC	4860
OY	4861	CCCTGCTGTGATCGGAAGAGAGGGAATATGTTATAGTTGTGACGACAGGCCAAGTT	4920
Db	4861	CCCTGCTGTGATCGGAAGAGAGGGAATATGTTATAGTTGTGACGACAGGCCAAGTT	4920
OY	4921	CAATATTTTCTGGGCAAAAACCTTCTTCAAAAATTAATGTAATCTTATTTATCAATGA	4980
Db	4921	CAATATTTTCTGGGCAAAAACCTTCTTCAAAAATTAATGTAATCTTATTTATCAATGA	4980
OY	4981	ATTACCTTGGAAATGACACCGCTCAACTGTGTCATGCGCTAAATGAAGGAATTTTA	5040
Db	4981	ATTACCTTGGAAATGACACCGCTCAACTGTGTCATGCGCTAAATGAAGGAATTTTA	5040
OY	5041	TAGTCTCTTAATAGGGGTGTACTGCAAGACCTTTGAACACTTTTCCAGAGATAGGATAT	5100
Db	5041	TAGTCTCTTAATAGGGGTGTACTGCAAGACCTTTGAACACTTTTCCAGAGATAGGATAT	5100
OY	5101	TTAATGTCATGCTTGGCGTTGTGCTATGCGACCTTTCCTTGTGAAGTCGTGTCTGCG	5160
Db	5101	TTAATGTCATGCTTGGCGTTGTGCTATGCGACCTTTCCTTGTGAAGTCGTGTCTGCG	5160
OY	5161	CCAATGACCCCTTGGCCTGTGAGCCGAGATGTGACCCCTGCTAAAGGCCAAAGAGAGG	5220
Db	5161	CCAATGACCCCTTGGCCTGTGAGCCGAGATGTGACCCCTGCTAAAGGCCAAAGAGAGG	5220
OY	5221	CTGGGCTTCTTCCCTCACTGTAAGAGCCCTTAATTTGAATTCACGTGTGAGGCCCTAGC	5280
Db	5221	CTGGGCTTCTTCCCTCACTGTAAGAGCCCTTAATTTGAATTCACGTGTGAGGCCCTAGC	5280
OY	5281	CTCTCATTTCTGACATATCCCAACCTCCACGCCCTTCCACAGACAGACTAGTGCCTGCG	5340
Db	5281	CTCTCATTTCTGACATATCCCAACCTCCACGCCCTTCCACAGACAGACTAGTGCCTGCG	5340
OY	5341	ATTTCACCCCAAGGTGGGAATGTGGCTTCCTTATAGCGGCTACTGTGACACATCACCGCAT	5400
Db	5341	ATTTCACCCCAAGGTGGGAATGTGGCTTCCTTATAGCGGCTACTGTGACACATCACCGCAT	5400

QY	5401	CACTGTGCTTCGACAGACACACAGTGGCCATTTTCTTCAACTGAGGCTCAAAACTCC	5460
Db	5401	CACTGTGCTTCGACAGACACACAGTGGCCATTTTCTTCAACTGAGGCTCAAAACTCC	5460
QY	5461	TGACACAGTTGCTGGGCTCCTGAGACAGATTTTCCGTGGAGCTGTGGCTCAGTGAAGGGGC	5520
Db	5461	TGACACAGTTGCTGGGCTCCTGAGACAGATTTTCCGTGGAGCTGTGGCTCAGTGAAGGGGC	5520
QY	5521	CCAGCCTGAGGAACCTGGCTCTTTTCTTTTAAAGCCACAGGCCCACTTACATTAACATT	5580
Db	5521	CCAGCCTGAGGAACCTGGCTCTTTTCTTTTAAAGCCACAGGCCCACTTACATTAACATT	5580
QY	5581	TCAGGGTCACTGGAAACAGTGAAGTCCATTTTGTGAAGCTACTGTGCATGCAGCCCACT	5640
Db	5581	TCAGGGTCACTGGAAACAGTGAAGTCCATTTTGTGAAGCTACTGTGCATGCAGCCCACT	5640
QY	5641	GCTCATCCACGTGGTCTGGCATGCTTACGAGGAAGGCCACGCCCATGTGCAGACTGTGCTCT	5700
Db	5641	GCTCATCCACGTGGTCTGGCATGCTTACGAGGAAGGCCACGCCCATGTGCAGACTGTGCTCT	5700
QY	5701	AATGCTGTGTTATTTGACAGAGAAGGAAAGTCTCAAGGAAGAAGTCAACTGGGACAGCA	5760
Db	5701	AATGCTGTGTTATTTGACAGAGAAGGAAAGTCTCAAGGAAGAAGTCAACTGGGACAGCA	5760
QY	5761	CAAGCCACCGGACATGGCTTTGTTAAAGTTAGCAGACTGGTGTGTGTGATCTGCAGT	5820
Db	5761	CAAGCCACCGGACATGGCTTTGTTAAAGTTAGCAGACTGGTGTGTGTGATCTGCAGT	5820
QY	5821	GCTTCACAGGAAATATTTATTTATTCATGTGCAGATCTTTTAACTGGCATTTTATCATTTTC	5880
Db	5821	GCTTCACAGGAAATATTTATTTATTCATGTGCAGATCTTTTAACTGGCATTTTATCATTTTC	5880
QY	5881	CTGTGCTTTAAATTAACAATTTACCAAAAAACAAGTATCAAGTCTGTTAACTGCTCGG	5940
Db	5881	CTGTGCTTTAAATTAACAATTTACCAAAAAACAAGTATCAAGTCTGTTAACTGCTCGG	5940
QY	5941	CTACTTGTCCCTGCTGTACAGTAGAGGCCCGGTTTCCAGTTGTTGACTGTGACAGGCTC	6000
Db	5941	CTACTTGTCCCTGCTGTACAGTAGAGGCCCGGTTTCCAGTTGTTGACTGTGACAGGCTC	6000
QY	6001	AGCATGGGCTGACGAGATGCTGTCTTAATTTTGTGATATACAGAAAGCCAGGCTTTGGG	6060
Db	6001	AGCATGGGCTGACGAGATGCTGTCTTAATTTTGTGATATACAGAAAGCCAGGCTTTGGG	6060
QY	6061	ATACAGTTCTTTTCTCTCTTCATTTATGACCGGACACTGTGTAAGCAGATGTTTTGTCC	6120
Db	6061	ATACAGTTCTTTTCTCTCTCTTCATTTATGACCGGACACTGTGTAAGCAGATGTTTTGTCC	6120
QY	6121	GGAATTAATAATATATAGTCTTGGAGCTTCGCCCAAAAAAAAAAAAA 6163	
Db	6121	GGAATTAATAATATATAGTCTTGGAGCTTCGCCCAAAAAAAAAAAAA 6163	
RESULT 2			
US-09-899-569A-1			
Sequence 1, Application US/09899569A			
Patent No. US20020142003A1			
GENERAL INFORMATION:			
APPLICANT: NO. US20020142003Albert Schweitzer			
APPLICANT: Marwa Scherl-Mostlager			
APPLICANT: Wolfgang Sommerhuber			
APPLICANT: Roger Abscher			
TITLE OF INVENTION: Tumorssoziliertes Antigen (3345)			
FILE REFERENCE: 0652.2280001			
CURRENT APPLICATION NUMBER: US/09/899,569A			
CURRENT FILING DATE: 2001-07-06			
PRIOR APPLICATION NUMBER: DE 100 33 080.0			
PRIOR FILING DATE: 2000-07-07			
PRIOR APPLICATION NUMBER: DE 101 19 294.0			
PRIOR FILING DATE: 2001-04-19			
PRIOR APPLICATION NUMBER: US 60/243,158			
PRIOR FILING DATE: 2000-10-25			
PRIOR APPLICATION NUMBER: US 60/2297,747			

```
; PRIOR FILING DATE: 2001-06-14
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: Patentl Ver. 2.1
; SEQ ID NO 1
; LENGTH: 5897
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5348)...(5348)
; OTHER INFORMATION: n is any nucleotide of a, t, g or c
; NAME/KEY: misc_feature
; LOCATION: (5360)...(5360)
; OTHER INFORMATION: n is any nucleotide of a, t, g or c
; NAME/KEY: misc_feature
; LOCATION: (5425)...(5425)
; OTHER INFORMATION: n is any nucleotide of a, t, g or c
; NAME/KEY: misc_feature
; LOCATION: (5438)...(5438)
; OTHER INFORMATION: n is any nucleotide of a, t, g or c
; NAME/KEY: 5'UTR
; LOCATION: (1)...(214)
; NAME/KEY: CDS
; LOCATION: (215)...(2464)
; NAME/KEY: 3'UTR
; LOCATION: (2465)...(5897)
; US-09-899-569A-1
```

```
Query Match          95.0%; Score 5854.2; DB 10; Length 5897;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 5865; Conservative 2; Mismatches 5; Indels 1; Gaps 1;
```

```
OY 291 CCTGAACGCGGGGCTCTATCGCACTGCTAGGGGTTCTGCTGCGGGCGCGGCT 350
DB 25 CCGTAACCTCGGGGCTCTATCGCACTGCTAGGGGTTCTGCTGCGGGCGCGGCT 84
OY 351 GCGCGCGCGGGGCAAGAGCTTTGAGATTGCTCTGCACGAGAAAGCAATTCAGTTCT 410
DB 85 GCGCGCGCGGGGCAAGAGCTTTGAGATTGCTCTGCACGAGAAAGCAATTCAGTTCT 144
OY 411 CATTAAGCTGGGGACCCGACCTCTGCTGGCAAAACCCGTTACATGTCATTTCTAAAG 470
DB 145 CATTAAGCTGGGGACCCGACCTCTGCTGGCAAAACCCGTTACATGTCATTTCTAAAG 204
OY 471 ACATTAACCATGTTGTCATCAAGCTCTGAGAGAAAGATAGCTTTACCTTTAGCGCA 530
DB 205 ACATTAACCATGTTGTCATCAAGCTCTGAGAGAAAGATAGCTTTACCTTTAGCGCA 264
OY 531 GAGTCTGAGATCACTTTGTCATAGAGATCCAGAAAATATTTGACTGTATGTCAGGGCC 590
DB 265 GAGTCTGAGATCACTTTGTCATAGAGATCCAGAAAATATTTGACTGTATGTCAGGGCC 324
OY 591 ATGTCCTTTGGGAGGTTGACCTTCAGCCCTCGACATCGTTGTTGCCCTCAACAG 650
DB 325 ATGTCCTTTGGGAGGTTGACCTTCAGCCCTCGACATCGTTGTTGCCCTCAACAG 384
OY 651 AACCTTCATCTGAGATGTCAAAGCTCATTAAGCATCGGTTTAGAGCTGACAGTTTCCAT 710
DB 385 AACCTTCATCTGAGATGTCAAAGCTCATTAAGCATCGGTTTAGAGCTGACAGTTTCCAT 444
OY 711 CCCGCGCGGAGAGATCGATCGGCTCGGGGTGAGAGCGCCAGCGGAGTACACATCCAT 770
DB 445 CCCGCGCGGAGAGATCGATCGGCTCGGGGTGAGAGCGCCAGCGGAGTACACATCCAT 504
OY 771 CAGCGCGCGAATCGATCGCACCGTGGTCAAGATCGGAACCTTCTGAGAGAAATGGGCACTGT 830
DB 505 CAGCGCGCGAATCGATCGCACCGTGGTCAAGATCGGAACCTTCTGAGAGAAATGGGCACTGT 564
OY 831 GTCCCGGATCAAGATGCAAGAAAGAGTGAATAATGGCTTACACCTCCATGGTTCCACCC 890
DB 565 GTCCCGGATCAAGATGCAAGAAAGAGTGAATAATGGCTTACACCTCCATGGTTCCACCC 624
OY 891 CAGAAATGTCCTCGGCTTCAGCATTTGCAAAACGCTCATGTATAAAACGCTGTGCATCAT 950
```

```
DB 625 CAGAAATGTCCTCGGCTTCAGCATTTGCAAAACGCTCATGTATAAAACGCTGTGCATCAT 684
OY 951 CGAGTCTGTTGTTGAGGGTGAAGGCTCAGCAACCCGTGATGTCTGCCACTACCCAGAG 1010
DB 685 CGAGTCTGTTGTTGAGGGTGAAGGCTCAGCAACCCGTGATGTCTGCCACTACCCAGAG 744
OY 1011 CTTCCTGAGAGATGACCTCATGACGTGGGAGTTGTCGTTCTCGCACCTCGGGGCCAG 1070
DB 745 CTTCCTGAGAGATGACCTCATGACGTGGGAGTTGTCGTTCTCGCACCTCGGGGCCAG 804
OY 1071 CGTCTCTTCCTCAACTTCAACCTCTCCAACTGTGAGAGAAAGAGAGCGGGTTAATA 1130
DB 805 CGTCTCTTCCTCAACTTCAACCTCTCCAACTGTGAGAGAAAGAGAGCGGGTTAATA 864
OY 1131 CTACATCCGCGGGCTCAGCACCAACCCGAGGTTCGAAGCTGAGAGACAGAGCCTGG 1190
DB 865 CTACATCCGCGGGCTCAGCACCAACCCGAGGTTCGAAGCTGAGAGACAGAGCCTGG 924
OY 1191 GAACATGGCGGGGAACTTCAACTCTCTCTGCAAGGCTGTGACCAAGATGCCAAGTGC 1250
DB 925 GAACATGGCGGGGAACTTCAACTCTCTCTGCAAGGCTGTGACCAAGATGCCAAGTGC 984
OY 1251 AGGATCTCTCGGCTGACAGTCCAACTTTGGTCCAACTCCCAAAATGAAAGCAATTA 1310
DB 985 AGGATCTCTCGGCTGACAGTCCAACTTTGGTCCAACTCCCAAAATGAAAGCAATTA 1044
OY 1311 AATCTACGTTGTTGACTTGAATAGAGAGGAGCATGTCATCCACATGAGAGCGGC 1370
DB 1045 AATCTACGTTGTTGACTTGAATAGAGAGGAGCATGTCATCCACATGAGAGCGGC 1104
OY 1371 CGTCAACAGAGCGGCAAGTTTGTCCCTGGCTGTTCGTGTGTCTGATCTTCGACCTG 1430
DB 1105 CGTCAACAGAGCGGCAAGTTTGTCCCTGGCTGTTCGTGTGTCTGATCTTCGACCTG 1164
OY 1431 CAGTACCAACCTCACCCCTGACATCTGCTGCTCAAAACCAAAATCTCTCTTGTGATGA 1490
DB 1165 CAGTACCAACCTCACCCCTGACATCTGCTGCTCAAAACCAAAATCTCTCTTGTGATGA 1224
OY 1491 TCTGACAGCTCTGTGATGATGTAATGTGAAAAAACCATTAAGCTGCACAGACACCGTATG 1550
DB 1225 TCTGACAGCTCTGTGATGATGTAATGTGAAAAAACCATTAAGCTGCACAGACACCGTATG 1284
OY 1551 CCAAGGAATCTTACTCTACTCTCCAGGTGCCAGTGAATCTCCACCTGCTGTGAGCT 1610
DB 1285 CCAAGGAATCTTACTCTACTCTCCAGGTGCCAGTGAATCTCCACCTGCTGTGAGCT 1344
OY 1611 GCATGACTCTCTCTGGAAGCTGCTGTGCTGCCAAGAGACAGGCTGAGCTGTGCTG 1670
DB 1345 GCATGACTCTCTCTGGAAGCTGCTGTGCTGCCAAGAGACAGGCTGAGCTGTGCTGCTG 1404
OY 1671 AGCCCAAGAGCTGACAGACATACACAGAGAAGCCCTCAACACAGCTTACAGTACCT 1730
DB 1405 AGCCCAAGAGCTGACAGACATACACAGAGAAGCCCTCAACACAGCTTACAGTACCT 1464
OY 1731 CGTGGCAATGCCATTAACAGCAGGACCTGTACTTCTGCTCTGCTGCTGCTGCTGCTG 1790
DB 1465 CGTGGCAATGCCATTAACAGCAGGACCTGTACTTCTGCTCTGCTGCTGCTGCTGCTG 1524
OY 1791 TATCAAGGATGCAAGTGAAGAGAAATCTCGGGGAGCCCTTGGACCTTGGCCCCCAG 1850
DB 1525 TATCAAGGATGCAAGTGAAGAGAAATCTCGGGGAGCCCTTGGACCTTGGCCCCCAG 1584
OY 1851 CTTCCAAGAGAGGCTCCAGGAGGCTGAGAGGCTGCTTATTAATCTTATTTCAAGA 1910
DB 1585 CTTCCAAGAGAGGCTCCAGGAGGCTGAGAGGCTGCTTATTAATCTTATTTCAAGA 1644
OY 1911 GGAAGGCTTTTACAGGCTGACCCCTGACCAAAAGAGTCTTACTGAGAGACCCCAA 1970
DB 1645 GGAAGGCTTTTACAGGCTGACCCCTGACCAAAAGAGTCTTACTGAGAGACCCCAA 1704
OY 1971 CTGGGACCGGGGCTGCACTCCCTGACCTGTGCTGTGGAACATAGGCTGCCAGAGA 2030
```

Db 1705 CTGGAGCCGGGGCTGCCATCCCTCACCTCTGTGTCTCTGAAATCATGACGCTGCCAGAGA 1764
QY 2031 CCAAGTGGCCCTGCTGACTTCTTTAAAGAGCGGAGCGGGGTGGTCTGCCAGACAGGGCG 2090
Db 1765 CCAAGTGGCCCTGCTGACTTCTTTAAAGAGCGGAGCGGGGTGGTCTGCCAGACAGGGCG 1824
QY 2091 CGCATTTATGATCATCCAGAGAGCAGCGGACCCGGGCTGAGAGATCTTCAAGCTTGAGCA 2150
Db 1825 CGCATTTATGATCATCCAGAGAGCAGCGGACCCGGGCTGAGAGATCTTCAAGCTTGAGCA 1884
QY 2151 GGATGTCTCCCCAAGCCAGCTTCCACATCCAGCTTGGGTGCAACATCTCTACTG 2210
Db 1885 GGATGTCTCCCCAAGCCAGCTTCCACATCCAGCTTGGGTGCAACATCTCTACTG 1944
QY 2211 CAGCCCCAGAGCGGCGAAGCAGCTAGACCTGTCTTCTCGGTGACACTTACCCCAAGAC 2270
Db 1945 CAGCCCCAGAGCGGCGAAGCAGCTAGACCTGTCTTCTCGGTGACACTTACCCCAAGAC 2004
QY 2271 TGTGGACTTGACTGTCTCTCTCATCGACGGGTGGAGTGTGAGTCTTACTGCTGTGC 2330
Db 2005 TGTGGACTTGACTGTCTCTCTCATCGACGGGTGGAGTGTGAGTCTTACTGCTGTGC 2064
QY 2331 CCTGGGCTCATCTTGTGTGTGAAAAAAGAAAAAGAAAAAGCAAAAGGCGCCGC 2390
Db 2065 CCTGGGCTCATCTTGTGTGTGAAAAAAGAAAAAGAAAAAGCAAAAGGCGCCGC 2124
QY 2391 TGTGGGTATCTACATGGAACAATCAATCAATGAGATGCCAGGCGCAAAAGTTTCA 2450
Db 2125 TGTGGGTATCTACATGGAACAATCAATCAATGAGATGCC - AGGAGCGCAAAAGTTTCA 2183
QY 2451 GAAAGGCGGAAAGGACATGACTCCCATGTGTATGACGTATGACGAGGACACATGTGATA 2510
Db 2184 GAAAGGCGGAAAGGACATGACTCCCATGTGTATGACGTATGACGAGGACACATGTGATA 2243
QY 2511 TGGGCACTGTGCTCAGGATTCAGGGGCTCCTTCTGCAAGAGGTGGAGACCTTACG 2570
Db 2244 TGGGCACTGTGCTCAGGATTCAGGGGCTCCTTCTGCAAGAGGTGGAGACCTTACG 2303
QY 2571 GCGGTTCCAGGGGACCATGGGGGTCTGTCCCTCCCAACCCACATATGTCTCAAGGC 2630
Db 2304 GCGGTTCCAGGGGACCATGGGGGTCTGTCCCTCCCAACCCACATATGTCTCAAGGC 2363
QY 2631 CCCAATCTGCAAGTTGGCCACTGAGAGCCACTCTCGCTCCCTGAGTCTGAGAG 2690
Db 2364 CCCAATCTGCAAGTTGGCCACTGAGAGCCACTCTCGCTCCCTGAGTCTGAGAG 2423
QY 2691 TGAACGCTACACCTTCCCATCCCAACCAATGGGGATGTAAAGAGCAAGAGACAGCAT 2750
Db 2424 TGAACGCTACACCTTCTCCCATCCCAACCAATGGGGATGTAAAGAGCAAGAGACAGCAT 2483
QY 2751 TCCCTTACTGACACTCTCAGAGGCCATGAGCCAGCAAGAACTTATGATCCATTCAGAC 2810
Db 2484 TCCCTTACTGACACTCTCAGAGGCCATGAGCCAGCAAGAACTTATGATTCAGAC 2543
QY 2811 GCTTTGCTGATTTCAATAAGCAGGCGCTGAGACACCCGCTCGTGTCTTAACAGAAA 2870
Db 2344 GCTTTGCTGATTTCAATAAGCAGGCGCTGAGACACCCGCTCGTGTCTTAACAGAAA 2603
QY 2871 TCCCTAAAGAGGAAATTAACAGAAAGAACAGAGAGTTTTCGAGGACCGGCCAAC 2930
Db 2604 TCCCTAAAGAGGAAATTAACAGAAAGAACAGAGAGTTTTCGAGGACCGGCCAAC 2663
QY 2931 TTCACATTTGCTCAGTGAATCTCTAAGGCGCAAGACATTTGAAATATGATTAATTCACATC 2990
Db 2664 TTCACATTTGCTCAGTGAATCTCTAAGGCGCAAGACATTTGAAATATGATTAATTCACATC 2723
QY 2991 TGGATACATGCTACATGACATCTCTCTCAACTTAAAGCTGTGCGGTTAGCCAGCTG 3050
Db 2724 TGGATACATGCTACATGACATCTCTCTCAACTTAAAGCTGTGCGGTTAGCCAGCTG 2783
QY 3051 TAATGAGAGAGAGGCGCTGATCCCTAGCATAGGTTGACAGCAAGCCCTGGATTACG 3110
Db 2784 TAATGAGAGAGAGGCGCTGATCCCTAGCATAGGTTGACAGCAAGCCCTGGATTACG 2843

QY 3111 AGTGTAAACAGAGGCTTGGCCCTCTTCAGAGACAACATTCCAATTCGAAGAGCTTACCT 3170
Db 2844 AGTGTAAACAGAGGCTTGGCCCTCTTCAGAGACAACATTCCAATTCGAAGAGCTTACCT 2903
QY 3171 GAGTCCCTACTCTCACTGGGGTCCCGAGATGAAAAAGCAATGTGCTTTTATATAT 3230
Db 2904 GAGTCCCTACTCTCACTGGGGTCCCGAGATGAAAAAGCAATGTGCTTTTATATAT 2963
QY 3231 ATTTATTTGGTGGTCCGTGTATTTAAAGATCAAAATGATAACCACTAGCTTTTC 3290
Db 2964 ATTTATTTGGTGGTCCGTGTATTTAAAGATCAAAATGATAACCACTAGCTTTTC 3023
QY 3291 ACCTGACTAGTAAATCACTACTAATCTGATTTGATGCTGGTTGTACTTCTACTG 3350
Db 3024 ACCTGACTAGTAAATCACTACTAATCTGATTTGATGCTGGTTGTACTTCTACTG 3083
QY 3351 ACCGCTAGATTAACGCTGTGCTGTCCCGAGGAGGTGGGAATTAATTAACAATCTGCCA 3410
Db 3084 ACCGCTAGATTAACGCTGTGCTGTCCCGAGGAGGTGGGAATTAATTAACAATCTGCCA 3143
QY 3411 CCAGAAAAAGATGTGTGTTGAGCAGCATATGACATATCTGCTTGAATAGAGACTT 3470
Db 3144 CCAGAAAAAGATGTGTGTTGAGCAGCATATGACATATCTGCTTGAATAGAGACTT 3203
QY 3471 CCTGATTTCTGTAGGTGGTCTGTGTATCCATTTGTGAATTTCAATCTTGAATCCATT 3530
Db 3204 CCTGATTTCTGTAGGTGGTCTGTGTATCCATTTGTGAATTTCAATCTTGAATCCATT 3263
QY 3531 GTCCATATAGTCCTTAGCAATTAAGAAATTTTCCCAATTTCCATGTGCGGTCTCCATG 3590
Db 3264 GTCCATATAGTCCTTAGCAATTAAGAAATTTTCCCAATTTCCATGTGCGGTCTCCATG 3323
QY 3591 TGCAGCAATCTTGTGACATTTAAAGAGAAATTAAGAAATATCTCATCTCTTAAATG 3650
Db 3324 TGCAGCAATCTTGTGACATTTAAAGAGAAATTTAGAAATATCTCATCTCTTAAATG 3383
QY 3651 TTTAAATATATACCAACAGTGGCCCTCATTTAGTTTCTGTGCACTGCAACCCAT 3710
Db 3384 TTTAAATATATACCAACAGTGGCCCTCATTTAGTTTCTGTGCACTGCAACCCAT 3443
QY 3711 TACTTGTACTTAAAAACAACATTAAGTATATAGTCTGGGGATAGAAATTCCAAAAT 3770
Db 3444 TACTTGTACTTAAAAACAACATTAAGTATATAGTCTGGGGATAGAAATTCCAAAAT 3503
QY 3771 GGATGTCCCTGAATGAATAATCAAGGTGTGAGAGAGTGTGCTCTCTGAAGGCTTAG 3830
Db 3504 GGATGTCCCTGAATGAATAATCAAGGTGTGAGAGAGTGTGCTCTCTGAAGGCTTAG 3563
QY 3831 GGAGAAAGCGGTTCTTGTCCATTTCAAGCTTCTAGAGGCTGTGATCCCAAGGCTCA 3890
Db 3564 GGAGAAAGCGGTTCTTGTCCATTTCAAGCTTCTAGAGGCTGTGATCCCAAGGCTCA 3623
QY 3891 GTGGCTGTGTAACGCTTTTCTACATGGCATCATGTGACACTGGCCCTCCACTTCCCTC 3950
Db 3624 GTGGCTGTGTAACGCTTTTCTACATGGCATCATGTGACACTGGCCCTCCACTTCCCTC 3683
QY 3951 TTTGACTTACAAAGCCACAGGAAGATCCAGATTAATCTCTCATATTAAGATCTTCA 4010
Db 3684 TTTGACTTACAAAGCCACAGGAAGATCCAGATTAATCTCTCATATTAAGATCTTCA 3743
QY 4011 TCATCTGGAAGACCTTTTGCATGCAAGACAATAGCCAGAGTGGGATTAAGACC 4070
Db 3744 TCATCTGGAAGACCTTTTGCATGCAAGACAATAGCCAGAGTGGGATTAAGACC 3803
QY 4071 AGACATCTTTGGGGGTCTTATTTGCTTACCAACCTTCTGCACTGACTCCACA 4130
Db 3804 AGACATCTTTGGGGGTCTTATTTGCTTACCAACCTTCTGCACTGACTCCACA 3863
QY 4131 GGAGAGGCTCAAAATGATGTGGGCGCAGAGGATTTGTTAGCTTGGGACTCTAAC 4190
Db 3864 GGAGAGGCTCAAAATGATGTGGGCGCAGAGGATTTGTTAGCTTGGGACTCTAAC 3923


```
; APPLICANT: Smith, Carole Lynn
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.471C13
; CURRENT APPLICATION NUMBER: US/09/922.217
; CURRENT FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 362
; LENGTH: 544
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-922-217-362

Query Match
Best Local Similarity 99.8%; Score 542.4; DB 10; Length 544;
Matches 543; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3068 CCGAGTCACCTAGACATAGAGTTGCACAGCCCTGGATTACAGAGTTAAACAGAGCT 3127
DB 1 CCGAGTCACCTAGACATAGAGTTGCACAGCCCTGGATTACAGAGTTAAACAGAGCT 60
OY -3128 TGCCCTCTTCAGGACAAACAGTTCCTCAATTCAGAGAGCTACCTGAGTCCCTACTCTCAC 3187
DB 61 TGCCCTCTTCAGGACAAACAGTTCCTCAATTCAGAGAGCTACCTGAGTCCCTACTCTCAC 120
OY 3188 TGGGGTCCCGAGATGAAAAAGACAAATGTGCTTTTATTATTATTATTATTGGTGCTCT 3247
DB 121 TGGGGTCCCGAGATGAAAAAGACAAATGTGCTTTTATTATTATTATTATTGGTGCTCT 180
OY 3248 GTGTTATTTAAGAGATCAATGTATTAACACCTAGCTCTTTTCACCTGACTTATTAATA 3307
DB 181 GTGTTATTTAAGAGATCAATGTATTAACACCTAGCTCTTTTCACCTGACTTATTAATA 240
OY 3308 CTCATACTAATCGTTTGGATGTCCTGGTGTGATCTTCTACTGACCGCTAGATTAACGTG 3367
DB 241 CTCATACTAATCGTTTGGATGTCCTGGTGTGATCTTCTACTGACCGCTAGATTAACGTG 300
OY 3368 TGCCGTGCCCCAGGTGGTGGGAATTAATTCATCTGTCCACCAACCAAGAAATGTGTG 3427
DB 301 TGCCGTGCCCCAGGTGGTGGGAATTAATTCATCTGTCCACCAACCAAGAAATGTGTG 360
OY 3428 TGTGTGAGCAGCATGTGACATATCTGCTTGTATTAAGAGACTTCCCTGATTCCTAGTGC 3487
DB 361 TGTGTGAGCAGCATGTGACATATCTGCTTGTATTAAGAGACTTCCCTGATTCCTAGTGC 420
OY 3488 GTTCGTGTTATCCCATTTGTGAATTCATCTTGAATCCCATTTGCTCTATAGTCTAGCA 3547
DB 421 GTTCGTGTTATCCCATTTGTGAATTCATCTTGAATCCCATTTGCTCTATAGTCTAGCA 480
OY 3548 ATAGAGAAATTTCCCAATTCCTCAATGTCGAGTCCCTAGTCCAGCAGCAATCTTGAC 3607
DB 481 ATAGAGAAATTTCCCAATTCCTCAATGTCGAGTCCCTAGTCCAGCAGCAATCTTGAC 540
OY 3608 ATTT 3611
DB 541 ATTT 544
```

```
; FILE REFERENCE: 210121.471C12
; CURRENT APPLICATION NUMBER: US/09/833.263
; CURRENT FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 1093
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 362
; LENGTH: 544
; TYPE: DNA
; ORGANISM: Homo sapien
; US-09-833-263-362

Query Match
Best Local Similarity 99.8%; Score 542.4; DB 10; Length 544;
Matches 543; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3068 CCGAGTCACCTAGACATAGAGTTGCACAGCCCTGGATTACAGAGTTAAACAGAGCT 3127
DB 1 CCGAGTCACCTAGACATAGAGTTGCACAGCCCTGGATTACAGAGTTAAACAGAGCT 60
OY 3128 TGCCCTCTTCAGGACAAACAGTTCCTCAATTCAGAGAGCTACCTGAGTCCCTACTCTCAC 3187
DB 61 TGCCCTCTTCAGGACAAACAGTTCCTCAATTCAGAGAGCTACCTGAGTCCCTACTCTCAC 120
OY 3188 TGGGGTCCCGAGATGAAAAAGACAAATGTGCTTTTATTATTATTATTATTGGTGCTCT 3247
DB 121 TGGGGTCCCGAGATGAAAAAGACAAATGTGCTTTTATTATTATTATTATTGGTGCTCT 180
OY 3248 GTGTTATTTAAGAGATCAATGTATTAACACCTAGCTCTTTTCACCTGACTTATTAATA 3307
DB 181 GTGTTATTTAAGAGATCAATGTATTAACACCTAGCTCTTTTCACCTGACTTATTAATA 240
OY 3308 CTCATACTAATCGTTTGGATGTCCTGGTGTGATCTTCTACTGACCGCTAGATTAACGTG 3367
DB 241 CTCATACTAATCGTTTGGATGTCCTGGTGTGATCTTCTACTGACCGCTAGATTAACGTG 300
OY 3368 TGCCGTGCCCCAGGTGGTGGGAATTAATTCATCTGTCCACCAACCAAGAAATGTGTG 3427
DB 301 TGCCGTGCCCCAGGTGGTGGGAATTAATTCATCTGTCCACCAACCAAGAAATGTGTG 360
OY 3428 TGTGTGAGCAGCATGTGACATATCTGCTTGTATTAAGAGACTTCCCTGATTCCTAGTGC 3487
DB 361 TGTGTGAGCAGCATGTGACATATCTGCTTGTATTAAGAGACTTCCCTGATTCCTAGTGC 420
OY 3488 GTTCGTGTTATCCCATTTGTGAATTCATCTTGAATCCCATTTGCTCTATAGTCTAGCA 3547
DB 421 GTTCGTGTTATCCCATTTGTGAATTCATCTTGAATCCCATTTGCTCTATAGTCTAGCA 480
OY 3548 ATAGAGAAATTTCCCAATTCCTCAATGTCGAGTCCCTAGTCCAGCAGCAATCTTGAC 3607
DB 481 ATAGAGAAATTTCCCAATTCCTCAATGTCGAGTCCCTAGTCCAGCAGCAATCTTGAC 540
OY 3608 ATTT 3611
DB 541 ATTT 544

RESULT 5
US-09-815-343-1335/C
; Sequence 1335, Application US/09815343
; Patient No. US20010055596A1
; GENERAL INFORMATION:
; APPLICANT: Meagher, Madeleine
; APPLICANT: Xu, Jiangchun
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.504
; CURRENT APPLICATION NUMBER: US/09/815.343
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1335
; LENGTH: 541
```


Tue Nov 12 18:03:25 2002

us-09-899-569a-3.mpb

Page 14

Job time : 755 secs

...

Qy	1883	CGGTCCTCTTAATACCTTAATTTCAAAGGAGGAGCGTTTACAGGTGACCCCTGACACAA	1942
Db	1741	TAGTCTCTATACACCCGATTTTCAAAGGAGGAGCATTTACAGGTGACCCGACACAA	1800
Qy	1943	AAACCAAGGTCTACCTGAGGAGCCCACTAGGAGCCGGGCGCTGCATCTCCATCCTGTG	2002
Db	1801	AAACCAAGGTCTACCTAGATATCACCACCTAGGAGCCGGGCGCTGCCTCTCTCTAG	1860
Qy	2003	TGTCCTGGACATCAGCGGTGCCAGAGACCAAGTGGCCCTGCATCTTTTAAGAGC	2062
Db	1861	TGTCCTGGACATCAGTGTGCCCTAGCAACCAAGTGGCTGTCTGACCGCTTTGAAGAGC	1920
Qy	2063	GGAGCGGGGTGTGTGCCAGAGCGGGCGCGCATTCATCATATCCAGAGAGCAGCGACCC	2122
Db	1921	GTTCCTGGTGTGCCCTGCAGTCAGGCGCGTGTATATATATATCCAGAGAGCAGCTCC	1980
Qy	2123	GGCGTGAAGAGATCTTCAGCCCTGGACGAGATGTGTCTCCCAAGCCAGCTTCCACATC	2182
Db	1981	GGCGAGAGAGATCTTCAGCTTGGAGGAGGAGAGTCTGTCTTAAGCCAGTTTCCACATC	2040
Qy	2183	ACAGCTTCTGGGTCAACATCTCTTAATCTGACAGCCCAAGAGCGGACAGCCTAGACCTGC	2242
Db	2041	ACAGCTTCTGGGTTCATCTTCATTCACATTCACCCCATTAAGGCAAGCAGCTAGATTGCG	2100
Qy	2243	TCTTCTGCGTACACTTATCCCAAGGAGCTGTGGATTTAGTGTATCTCTCATTCGAGGG	2302
Db	2101	TCTTCTGCGTACACTTATCCCGAGAGCTGTAGATTGCGTGG--TGTCTATGTGTGAG	2157
Qy	2303	TGGGAGTGGAGTCTTACTGTCTGTCTGCCCTCGGGCTCATATTTGCTGTGAAAAA	2362
Db	2158	CAGGAGTGGAGTCTGTCTGTCTGTGGTGGCTGTGCTGATCATCTGTCTGTGAAAAA	2217
Qy	2363	AGAAAAA 2369	
Db	2218	AAAAAAA 2224	
RESULT 2			
LOCUS	AB033417	933 bp	mRNA linear EST 18-Oct-1999
DEFINITION	AB033417 Homo sapiens skov Homo sapiens cDNA, mRNA sequence.		
ACCESSION	AB033417		
VERSION	AB033417.1	GI:6062545	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
AUTHORS	1 (bases 1 to 933)		
TITLE	Zhou Y., Kato H. and Wake, N.		
JOURNAL	Biological function unknown		
COMMENT	Unpublished (1999)		
	Contact: Zhou Yong		
	Department of Reproductive Physiology and Endocrinology		
	Medical Institute of Bioregulation, Kyushu University		
	Tsurunihara 4546, Bepu, Oita 874-0838, Japan		
	Tel: 0977-27-1660		
	Fax: 0977-27-1661		
	Email: yzhounetsu@uml.bepu.kyushu-u.ac.jp.		
FEATURES			
source	Location/Qualifiers		
	1..933		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone_lhb="Homo sapiens skov"		
	/cell_type="ovarian cancer"		
	/cell_line="skov"		
	/note="Buessow, K. et al., 1998, Nucleic Acids Research		
	26(21): 5007-5008"		
BASE COUNT	265 a 251 c 219 g 198 t		
ORIGIN			
Query Match	14.7%; Score 905.8; DB 9; Length 933;		

[illegible]

ACCESSION	AM063026
VERSION	AM063026.1
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
TITLE	1 (bases 1 to 933)
JOURNAL	Zhou,Y., Kato,H. and Wake,N.
COMMENT	A function unknown human partial mRNA, 3' region Unpublished (1999) Contact: Zhou Yong Department of Reproductive Physiology and Endocrinology Medical Institute of Bioregulation, Kyushu University Tsushima-hara 4546, Beppu, Oita 874-0838, Japan Tel: 0977-27-1660 Fax: 0977-27-1661 Email: yzhout@surumt.beppu.kyushu-u.ac.jp.
FEATURES	Location/Qualifiers
source	1..933
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="EST392"
	/clone_lib="human ovarian cancer CDNA"
	/tissue_type="ovarian cancer"
	/sex="female"
	/cell_line="skov"
	/note="Organ: ovary"
BASE COUNT	265 a 251 c 219 g 198 t
ORIGIN	
Query Match	14.7% Score 905.8; DB 10; Length 933;
Best Local Similarity	99.7%; Pred. No. 7.4e-115;
Matches 918; Conservative	0; Mismatches 2; Indels 1; Gaps 1.
OY	2315 TCTTTCGTGCTGCTCCCGCGGCATCATTTGGCTGTGAAGAAAGAAGAAAGA 2374
DB	1 TCTTACTGCTGCTCCTCGCGGCATCATTTGGCTGTGAAGAAAGAAGAAAGA 60
OY	2375 CAACAAGGGCCCCGCTGTGGGTATCTCAATGGCAATCAATACTAGATCCGAGGC 2434
DB	61 CAAACAAGGGCCCCGCTGTGGGTATCTCAATGGCAATCAATACTAGATCCGAGGC 120
OY	2435 AGCCAAAAAAGTTTAGAAAAGGGGCCAAAGAGCAATGATCCTCATGTGTATGATGATCG 2494
DB	121 AGCCAAAAAAGTTTAGAAAAGGGGCCAAAGAGCAATGATCCTCATGTGTATGATGATCG 180
OY	2495 AGGACACCATTGTAATGGGCACTGTGATACAGATTTCCAGGGGCTTCCTGCAAGCCAG 2554
DB	181 AGGACACCATTGTAATGGGCACTGTGATACAGATTTCCAGGGGCTTCCTGCAAGCCAG 240.
OY	2555 AGGTGGACACCTTACCGGCGTTCCAGGGCACCATGGGGTGCTCTCCCTCCCACCCA 2614
DB	241 AGGTGGACACCTTACCGGCGTTCCAGGGCACCATGGGGTGCTCTCCCTCCCACCCA 300
OY	2615 CCATATGCTCCAGGGCCCCAACCTGCAAAAGTTGGCACTGAGGAGCCACCTCTCGCTCCC 2674
DB	301 CCATATGCTCCAGGGCCCCAACCTGCAAAAGTTGGCACTGAGGAGCCACCTCTCGCTCCC 360
OY	2675 CTCCTGAGTCTGAGAGTGAACCGTACACCTTCTCCATCTCCAAACAATGGGGATGTAAGA 2734
DB	361 CTCCTGAGTCTGAGAGTGAACCGTACACCTTCTCCATCTCCAAACAATGGGGATGTAAGA 420
OY	2735 GCAAAGGACACAGACATTTCCCTTACTGAAACACTCAGAGAGCCATGAGAGCAGAGATATAC 2794
DB	421 GCAAAGGACACAGACATTTCCCTTACTGAAACACTCAGAGAGCCATGAGAGCAGAGATATAC 480
OY	2795 TTGATCCATTCACAGAGCGTTTGCTGAGTTTCAATAAGCAGGGCACTGAGACACCGCTCG 2854
DB	481 TTGATCCATTCACAGAGCGTTTGCTGAGTTTCAATAAGCAGGGCACTGAGACACCGCTCG 540
OY	2855 TGTTCTTACCAAGAATCTTAAGAAGGAAATTATACAGAAAGCAACGAGAGGTTTT 2914

DB	541	TGTTCTTAACCGAAGAAATCTCTAAGAGAGAGAAATTTATACAGAGAGAAACAGACAGAGAGTTT	600
QY	2915	CCTGACACACGCCCAACTTACATTGCTCAGTGGACTCATTTAGGGCAACATTTGAAA	2974
Db	601	CCTGGACACACGCCCAACTTACATTGCTCAGTGGACTCATTTAGGGCAACATTTGAAA	660
QY	2975	ATGATGAATTTCCAAATCTGGATTCAGTCAATGACAGCTCATGTCCTCTCAACTTAGGCTGT	3034
Db	661	ATGATGAATTTCCAAATCTGGATTCAGTCAATGACAGCTCATGTCCTCTCAACTTAGGCTGT	720
QY	3035	GGGTTTACCCAGCCCTGTATATGAGAGAGAGAGAGCCCTGATGCACCTACATAGATAGGTTGCG	3094
Db	721	GGGTTTACCCAGCCCTGTATATGAGAGAGAGAGAGCCCTGATGCACCTACATAGATAGGTTGCG	780
QY	3095	CAGGCCCTGGATTCAGAGTGTAAACAGAGGCTTCCCTTTACAGACACAGTTTCCAA	3154
Db	781	CGAGCCCTGGATTCAGAGTGTAAACAGAGGCTTCCCTTTACAGACACAGTTTCCAA	840
QY	3155	TCCAGAGAGCCCTGAGTGGTCCCTACCTGCTGAGGCTCCAGATGAAGACACAT	3214
Db	841	TCCAGAGAGCCCTGAGTGGTCCCTACCTGCTGAGGCTCCAGATGAAGACACAT	899
QY	3215	GTGCTTTTATTTATTTATTTA	3235
Db	900	GTGCTTTTATTTATTTATTTA	920
RESULT 4			
BO962997		928 bp	mrna linear EST 21-AUG-2002
LOCUS	AGNCNCURT_8778970	NIH_MGC_42	Homo sapiens cdna clone IMAGE:6378668
DEFINITION	5', mRNA sequence.		
ACCESSION	BO962997		
VERSION	BO962997.1	GI:22378463	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
TITLE	1 (bases 1 to 928)		
JOURNAL	NIH-MGC http://mgs.nci.nih.gov/.		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgapbs-remail.nih.gov		
	Tissue Procurement: ATCC		
	CDNA Library Preparation: Rubin Laboratory		
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)		
	DNA Sequencing by: Agencourt Bioscience Corporation		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LNL at:		
	http://image.lnl.gov		
	plate: LCM2564 row: 9 column: 21		
	High quality sequence stop: 589.		
FEATURES			
Source	Location/Qualifiers		
	1..928		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:6378668"		
	/clone_1ib="NIH_MGC_42"		
	/tissue_type="epithelioid carcinoma cell line"		
	/lab_host="DH10B (phage-resistant)"		
	/note="Organ: pancreas; Vector: pOTR7; site_1: XhoI;		
	site_2: EcoRI; CDNA made by oligo-dT priming.		
	directionally cloned into EcoRI/XhoI sites using the		
	following 5' adaptor: GGCACGAG(G). Size-selected >500bp		
	for average insert size 1.8kb. Library constructed by Ling		
	Hong in the laboratory of Gerald M. Rubin (University of		
	California, Berkeley) using ZAP-CDNA synthesis kit		
	(Stratagene) and Superscript II RT (Life Technologies).		
	Note: this is a NIH_MGC Library. I"		
BASE COUNT	219 a	257 c	238 g
ORIGIN			214 t

Db 361 GTGGCCATTTCTTCACTGAGGGCTCAAAACTCTGACAGTCTGCTCTGAGA 420
 QY 5485 CCAGTATTTCTCGAGCTGTGCTCACTGAGAGGGGCCAGCTGAGAACCTGCTCTT 5544
 Db 421 CCACTATTTCTCGAGCTGTGCTCACTGAGAGGGGCCAGCTGAGAACCTGCTCTT 480
 QY 5545 TTCTTTAAAGCCCGCCCACTTACATTAACATTTTCAGGCTGCTGAGAACGTAAG 5604
 Db 481 TTCTTTAAAGCCCGCCCACTTACATTAACATTTTCAGGCTGCTGAGAACGTAAG 540
 QY 5605 TGCATTTGTTGAAGCTTACTGATGCTGAGCCCACTGCTCATCCAGCTGCTGCTGAG 5664
 Db 541 TGCATTTGTTGAAGCTTACTGATGCTGAGCCCACTGCTCATCCAGCTGCTGCTGAG 600
 QY 5665 CTACAGGAGAGGCGCAGCGATGACAGAGCTGCTCTAATGCTGTGTCATTGACAGAG 5724
 Db 601 CTACAGGAGAGGCGCAGCGATGACAGAGCTGCTCTAATGCTGTGTCATTGACAGAG 660
 QY 5725 GGAAGGCTTCAGAGAGATCACTGAGTGGACAGACAGACAGCCCACTGAGCTTGG 5784
 Db 661 GGAAGGCTTCAGAGAGATCACTGAGTGGACAGACAGACAGCCCACTGAGCTTGG 720
 QY 5785 TAAAGTTACAGAGCTGTGTGATGCTGAGTCTGCTGCTGAGAAATTAATTATCA 5844
 Db 721 TAAAGTTACAGAGCTGTGTGATGCTGAGTCTGCTGCTGAGAAATTAATTATCA 780
 QY 5845 TTGCAGATCTTTTAAAGTGG-CATTTTATTCATTTCTGCTGCTGCTGCTGCTGCTG 5902
 Db 781 TTGCAGATCTTTTAAAGTGGCCATTTTATTCATTTCTGCTGCTGCTGCTGCTGCTG 840
 QY 5903 TACCAAAAACAGTATCA 5922
 Db 841 TACCAAAAACAGTATCA 860

RESULT 6
 BO691792 1008 bp mRNA linear EST 15-JUL-2002
 LOCUS BO691792
 DEFINITION AGENCOURT_8034267 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6208223
 5', mRNA sequence.
 ACCESSION BO691792
 VERSION BO691792.1 GI:21817108
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE NIH-MGC http://mgi.nci.nih.gov/
 1 (bases 1 to 1008)
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished (1999)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: c9abs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LINL at:
 http://image.linl.gov
 Plate: LHCW365 row: 1 column: 24
 High quality sequence start: 27
 High quality sequence stop: 744.
 Location/Qualifiers
 1..1008
 FEATURES
 source
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:6208223"
 /clone_lib="NIH_MGC_110"
 /tissue_type="ductal carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: pancreas; Vector: pORF7; Site_1: XhoI;
 Site_2: EcoRI; cDNA made by oligo-dT priming."

BASE COUNT 238 a 238 c 224 g 246 t 62 others
 ORIGIN
 Query Match 13.1%; Score 810.4; DB 14; Length 1008;
 Best Local Similarity 98.6%; Pred. No. 1e-102;
 Matches 847; Conservative 0; Mismatches 9; Indels 3; Gaps 3;
 Note: this is a NIH-MGC library.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCAAGAG(G). Library constructed by
 Ling Hong in the laboratory of Gerald M. Rubin (University
 of California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 QY 5289 CTGCAATTCCTCCCACTCCAGCCCTTCCCAAGAGAGTGGCTGATTCAC 5348
 Db 69 CTCGACATTCCTCCCACTCCAGCCCTTCCCAAGAGAGTGGCTGATTCAC 128
 QY 5349 CAAGGTGGATTTGGCTTCTTCTTAAAGGCTGCTGCTGCTGCTGCTGCTG 5408
 Db 129 CAAGGTGGATTTGGCTTCTTCTTAAAGGCTGCTGCTGCTGCTGCTGCTG 188
 QY 5409 CTGCAAGAGACACAGCTGGCCATTTCTTCACTAGAGGCTCAAACTCTGACAG 5468
 Db 189 CTGCAAGAGACACAGCTGGCCATTTCTTCACTAGAGGCTCAAACTCTGACAG 248
 QY 5469 TTGCTGCTCTCTGAGACAGTATTTCTTGAAGCTGCTGCTGCTGCTGCTG 5528
 Db 249 TTGCTGCTCTCTGAGACAGTATTTCTTGAAGCTGCTGCTGCTGCTGCTG 308
 QY 5529 AGGACCTGCTGCTTCTTCTTAAAGGCTGCTGCTGCTGCTGCTGCTGCTG 5588
 Db 309 AGGACCTGCTGCTTCTTCTTAAAGGCTGCTGCTGCTGCTGCTGCTGCTG 368
 QY 5589 ACTGGAACAGTGAAGTCCATTTGTTGAAGCTGCTGCTGCTGCTGCTGCTG 5648
 Db 369 ACTGGAACAGTGAAGTCCATTTGTTGAAGCTGCTGCTGCTGCTGCTGCTG 428
 QY 5649 ACGTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5708
 Db 429 ACGTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 488
 QY 5709 GGTGATTCAGACAGAGGAGGAGGCTGCAAGAGAGTCACTGGGAGACAGACCA 5768
 Db 489 GGTGATTCAGACAGAGGAGGAGGCTGCAAGAGAGTCACTGGGAGACAGACCA 548
 QY 5769 CCGGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5828
 Db 549 CCGGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 608
 QY 5829 GGAATATATTTATTCATTCAGATACATTTTATGCTGCTGCTGCTGCTGCT 5888
 Db 609 GGAATATATTTATTCATTCAGATACATTTTATGCTGCTGCTGCTGCTGCT 668
 QY 5889 TAAATTAACAAGTACCAAAAAACAAGTACCAAGTACCAAGTACCAAGTACCA 5948
 Db 669 TAAATTAACAAGTACCAAAAAACAAGTACCAAGTACCAAGTACCAAGTACCA 728
 QY 5949 CCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6008
 Db 729 CCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 788
 QY 6009 CTCAGCAGATGCTGCTTAAATTTGTGATATACAGAAAGCCAGGCTTGGATACAG 6067
 Db 789 CTCAGCAGATGCTGCTTAAATTTGTGATATACAGAAAGCCAGGCTTGGATACAG 848
 QY 6068 TTCTTTCTCTTCAATTTGATGCGG-TGCACGTGTGTAAGAGATGTTTGTGCGG-AA 6125
 Db 849 NTCTTTCTCTTCAATTTGATGCGG-TGCACGTGTGTAAGAGATGTTTGTGCGG-AA 908
 QY 6126 TAAATAATATATGCTTGA 6144
 Db 909 TAAATAATATGCTTGA 927

RESULT 7
LOCUS B0951236 984 bp mRNA linear EST 21-AUG-2002
DEFINITION AGENCOURT.8817950 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:642238
5', mRNA sequence.
ACCESSION B0951236
VERSION B0951236.1 GI:22366714
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE NIH-MGC <http://mgs.nci.nih.gov/>.
1 (bases 1 to 984)
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: DCTD/DP/GenDAR
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LICM2598 row: 0 column: 07
High quality sequence start: 6
High quality sequence stop: 719.
Location/Qualifiers
1..984
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="642238"
/clone_lib="NIH_MGC_18"
/tissue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pORF7; Site:1: XhoI; Site:2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGAG(G). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC library."

BASE COUNT 238 a 301 c 247 g 197 t 1 others
ORIGIN

Query Match 12.9% Score 794.4; DB 14; Length 984;
Best Local Similarity 96.7%; Pred. No. 1.6e-100;
Matches 832; Conservative 0; Mismatches 26; Indels 2; Gaps 2;

QY 951 CGAGTGTGTGTTGAGGGGTGAGGCTCAGCAACCTGATGTCGCCAATACCCAGAGG 1010
DB 18 CGAGGCTGTGTTGAGGGGTGAGGCTCAGCAACCTGATGTCGCCAATACCCAGAGG 77
QY 1011 CTTCCCTGAGAGTACTATGACGTGGGCGAGTTTGCTGCTCCGACACCTGGCGGCAG 1070
DB 78 CTTCCCTGAGAGTACTATGACGTGGGCGAGTTTGCTGCTCCGACACCTGGCGGCAG 137
QY 1071 CGTCTCTCTCTCACTCACTCACTCACTGAGAGAGAGAGAGAGGGGTTGAATA 1130
DB 138 CGTCTCTCTCTCACTCACTCACTCACTGAGAGAGAGAGAGAGGGGTTGAATA 137
QY 1131 CTACATCCGGGGCTCCACCAACCCCGAGGTTCACAGCTGGAGAGACAGCCTGG 1190
DB 198 CTACATCCGGGGCTCCACCAACCCCGAGGTTCACAGCTGGAGAGACAGCCTGG 257
QY 1191 GAACATGGGGGGGAACTTCAACTCTCTCTGCAAGGCTGGACCAAGATCCCAAGTCC 1250
DB 258 GAACATGGGGGGGAACTTCAACTCTCTCTGCAAGGCTGGACCAAGATCCCAAGTCC 317
QY 1251 AGGATCTCTCGGCTGCAAGTTTGTGTCACATCCACAAATGAAGCAATAA 1310

DB 318 AGGATCTCTCGGCTGCAAGTTTGTGTCACATCCACAAATGAAGCAATAA 377
QY 1311 AATCTACGTGTTGACTTGAATGATGAGCAGCAGCATCTACTACATGACCGACGCC 1370
DB 378 AATCTACGTGTTGACTTGAATGATGAGCAGCAGCATCTACTACATGACCGACGCC 437
QY 1371 CGTCAACAGAGCGGCAAGTTTGCCCTGGCTTTTCGTGCTGTAATCTGGACCTG 1430
DB 438 CGTCAACAGAGCGGCAAGTTTGCCCTGGCTTTTCGTGCTGTAATCTGGACCTG 497
QY 1431 CAGTAGCAACCTCACCCTGACATCTGGCTCCAAACACAAATCTCCTTTGTGATGA 1490
DB 498 CAGTAGCAACCTCACCCTGACATCTGGCTCCAAACACAAATCTCCTTTGTGATGA 557
QY 1491 TCTGACACGTCTGTGATGATGATGTAATGAAAAACCATTAAGTGCACAGACCGTACTG 1550
DB 558 TCTGACACGTCTGTGATGATGATGTAATGAAAAACCATTAAGTGCACAGACCGTACTG 617
QY 1551 CCAAGGAATCCTACTACTCCAGGAGGCCAGTACATCTCAGCTGCTGGAGCT 1610
DB 618 CCAAGGAATCCTACTACTCCAGGAGGCCAGTACATCTCAGCTGCTGGAGCT 677
QY 1611 GCATGACTTCTCTGGAAGCTGTGTGCCCCAAGAGACAGCTCAGCTGTGTGCC 1670
DB 678 GCATGACTTCTCTGGAAGCTGTGTGCCCCAAGAGACAGCTCAGCTGTGTGCC 737
QY 1671 AGCCCAAGACCTGACGACGATACACAGAGAAAGCCCTGCAACACGAC-TTACGTAAC 1729
DB 738 AGCCCAAGACCTGACGACGATACACAGAGAAAGCCCTGCAACACGATTAACCTAC 797
QY 1730 TCGTGGCCATGCGCATTCACAGCC-AGACCTGACTTGCGCTCTGCCGGGAGGC 1788
DB 798 TCGGCGGCGACTGACATCCAGCCAGAGACCTGACTTCGGCTCTTTGCCCCGGGA 857
QY 1789 TCTATCAAGCAGATCCAGGT 1808
DB 858 GAGCTCTATCAACCAAGAT 877

RESULT 8
LOCUS B0889198 871 bp mRNA linear EST 16-AUG-2002
DEFINITION AGENCOURT.8072366 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6084315
5', mRNA sequence.
ACCESSION B0889198
VERSION B0889198.1 GI:22281212
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE NIH-MGC <http://mgs.nci.nih.gov/>.
1 (bases 1 to 871)
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LICM2315 row: 0 column: 04
High quality sequence stop: 707.
Location/Qualifiers
1..871
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="6084315"
/clone_lib="NIH_MGC_110"

FEATURES
source

Db	Accession	LOCUS	DEFINITION	ORGANISM	Source
Db	361	GGCCATTGGACGCCAGAGAAATTAAC	851 bp	Human sapiens	High quality sequence stop: 699.
Db	2832	CAGGGCACTGAGACACCCGTC	Linear	Human sapiens	High quality sequence stop: 699.
Db	421	CAGGGCACTGAGACACCCGTC	Linear	Human sapiens	High quality sequence stop: 699.
Db	2892	CAGAGGAACACAGCAGGAGGTTTCT	Linear	Human sapiens	High quality sequence stop: 699.
Db	481	CAGAGGAACACAGCAGGAGGTTTCT	Linear	Human sapiens	High quality sequence stop: 699.
Db	2952	ATTCTTAAGGGCAGACATTTGAAAT	Linear	Human sapiens	High quality sequence stop: 699.
Db	541	ATTCTTAAGGGCAGACATTTGAAAT	Linear	Human sapiens	High quality sequence stop: 699.
Db	3012	ATGTCCTCTCAACCTTAAGGCTGT	Linear	Human sapiens	High quality sequence stop: 699.
Db	601	ATGTCCTCTCAACCTTAAGGCTGT	Linear	Human sapiens	High quality sequence stop: 699.
Db	3072	AGTACCTTAAGGCTGTGCGGTTAG	Linear	Human sapiens	High quality sequence stop: 699.
Db	661	AGTACCTTAAGGCTGTGCGGTTAG	Linear	Human sapiens	High quality sequence stop: 699.
Db	3132	CTCTTCAGAGACAGAGGTTTCCAT	Linear	Human sapiens	High quality sequence stop: 699.
Db	721	TCTTGAAGGACCCAGTTTCAATTT	Linear	Human sapiens	High quality sequence stop: 699.
Db	3191	GGTCCCGCAGAGTGAATAC	Linear	Human sapiens	High quality sequence stop: 699.
Db	781	GGGTCCTCCAGATTTGAAC	Linear	Human sapiens	High quality sequence stop: 699.

BASE COUNT	161 a	236 c	240 g	194 t	
ORIGIN	(Stratageme) and Superscript II RT (Life Technologies)."				
Query Match	11.5%; Score 709.4; DB 12; Length 851;				
Best Local Similarity	94.4%; Pred. No. 9.4e-89;				
Matches 779; Conservative	0; Mismatches 41; Indels 5; Gaps 4;				
OY	148 GGGCGGGGCTCGGGCGGGTCCGCCGCGCGAGGTGAGTGAACCCAGGCGGAGGCGACGT 207				
DB	2 GGGCGGGGCTCGGGCGGGTCCGCCGCGCGAGGTGAGTGAACCCAGGCGGAGGCGACGT 60				
OY	208 GCGCCGGGCTTGGGCGGCTGGGGCCGCGCTCCACCGTGGTTTTCCACCGAGCGG 267				
DB	61 GCGCCGGGCTTGGGCGGCTGGGGCCGCGCTCCACCGTGGTTTTCCACCGAGGCGG 120				
OY	268 AGGCGTCCCGGAGTCATGGCCGGCCGTGAACCTCGGGGGTCTCATATGCACTGTAGGGGT 327				
DB	121 AGGCGTCCCGGAGTCATGGCCGGCCGTGAACCTCGGGGGTCTCATATGCACTGTAGGGGT 180				
OY	328 CTGCGTGTGGGCGGCGCGCTGGCGGCGGGGAGAACTTTGAGATTGCTGTGCA 387				
DB	161 CTGCGTGTGGGCGGCGCGCTGGCGGCGGGGAGAACTTTGAGATTGCTGTGCA 240				
OY	388 CGAGAAGCACTTACAGTTCTCATTAAGCTGGGGACCCCGACCTGCTGGCAAAACC 447				
DB	241 CGAGAAGCACTTACAGTTCTCATTAAGCTGGGGACCCCGACCTGCTGGCAAAACC 300				
OY	448 TGTTCATCTGTCATTTCTTAAAGACATATTAACATGTTGTCCATGAAGCTGGAGAA 507				
DB	301 TGTTCATCTGTCATTTCTTAAAGACATATTAACATGTTGTCCATGAAGCTGGAGAA 360				
OY	508 ATAGCTTTTACCTTTAGCTGCGAGAGTCCGTGAGATTCACCTTTGTGATAGAGATCCAGAA 567				
DB	361 ATAGCTTTTACCTTTAGCTGCGAGAGTCCGTGAGATTCACCTTTGTGATAGAGATCCAGAA 420				
OY	568 AATATTTGACTGTATGTCAAGCCCATGTCTCTTTGGGGAGTTCAAGCTTCAAGCCCTGACA 627				
DB	421 AATATTTGACTGTATGTCAAGCCCATGTCTCTTTGGGGAGTTCAAGCTTCAAGCCCTGACA 480				
OY	628 TCGTTTGGCCCAACCCCTCAACAGAACTTTCATCTGGAGTGTCAAGCTCATTAAGAGCATC 687				
DB	481 TCGTTTGGCCCAACCCCTCAACAGAACTTTCATCTGGAGTGTCAAGCTCATTAAGAGCATC 540				
OY	688 GGTTTAGAGCTTCAGATTTTCCATCCCTGCGCTGAGGAGATGGTCCGGGTGAGAGCTGC 747				
DB	541 GGTTTAGAGCTTCAGATTTTCCATCCCTGCGCTGAGGAGATGGTCCGGGTGAGAGCTGC 600				
OY	748 CCAGACGGAGTCACTCACTCCATCCATCAAGCGGCCGAATCGATGCCACCGTGGTCAAGATCGGA 807				
DB	601 CCAGACGGAGTCACTCACTCCATCCATCAAGCGGCCGAATCGATGCCACCGTGGTCAAGATCGGA 660				
OY	808 ACCTTCTGCAGCAA-TGGACATGTGTCCC-GGATCAAGATGCAAGAAGS--AGTAAAAAT 863				
DB	661 ACCTTCTGCAGCAAATTTGGACATGTGTCCCAGGAGATCAAGATGCAAGAAGGAGTAAAAAT 720				
OY	864 GGCCTTACACCCCTCCATGGTTCCACCCCGGAATGTCTCCGGCTTCAGCATTTGCAAAACG 923				
DB	721 GGCCTTACACCTCCATGGTTCCACCCCGGAATGTCTCCGGCTTCAGCATTTGCAAAACG 780				
OY	924 CTCATCTATAAAGCTGTGTGCATCATCGAGTCTGTGTTGAAGG 968				
DB	781 GTCATTTATAAAGTTGTGTGCTATGCTTTGGGTTGAGGGTGAAG 825				
RESULT 11					
Bg119556					
LOCUS	Bg119556	779 bp	mRNA	linear	EST 30-JAN-2001
DEFINITION	603347259P1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4442127 5',				
ACCESSION	Bg119556				
VERSION	Bg119556.1				
KEYWORDS	EST. 12613062				

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 779)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov
Plate: LLM0214 row: f column: 16
High quality sequence stop: 700.

FEATURES
source
1. 779
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4442127"
/clone_1lb="NIH_MGC_90"
/issue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dt primed. Average insert size 1.7 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."

BASE COUNT 185 a 215 c 161 g 218 t

ORIGIN

Query Match 11.2%; Score 691.2; DB 12; Length 779;
Best Local Similarity 96.1%; Pred. No. 3.2e-86;
Matches 742; Conservative 0; Mismatches 23; Indels 7; Gaps 3;

QY 3977 ATCCAGATATATCTCCATCTAAAGATCCCTTCATCTCTGGAAGACCTTTGGCATG 4036
|||||
Db 1 ATCCAGATATATCTCCATCTAAAGATCCCTTCATCTCTGGAAGACCTTTGGCATG 60
|||||

QY 4037 CAAGACAACATAGCCACAGGTGGGATTTAGACACGACATCTTTGGGGTCTGTATTC 4096
|||||
Db 61 CAAGACAACATAGCCACAGGTGGGATTTAGACACGACATCTTTGGGGTCTGTATTC 120
|||||

QY 4097 TGCTTACACACTTCTGCGACACTCCACAGAGAGGCTCAAAATGATCTGGCGC 4156
|||||
Db 121 TGCTTACACACTTCTGCGACACTCCACAGAGAGGCTCAAAATGATCTGGCGC 180
|||||

QY 4157 ACAGGATGTTTTGTTAGTTGGGACTGTAACTT-AAAAAACCCAGATCAGAG 4215
|||||
Db 181 ACAGGATGTTTTGTTAGTTGGGACTGTAACTTAAAAAACCCAGATCAGAG 240
|||||

QY 4216 ATCTGCCATGCTGGGCTCACAATTCACCTAGCAACAACCTGGCTGAGCTGGGACCA 4275
|||||
Db 241 ATCTGCCATGCTGGGCTCACAATTCACCTAGCAACAACCTGGCTGAGCTGGGACCA 300
|||||

QY 4276 GCTGTGCTTTAGAGGGGTGCTCACTTCACTGACACACACCCCACTAGGCCCTA 4335
|||||
Db 301 GCTGTGCTTTAGAGGGGTGCTCACTTCACTGACACACACCCCACTAGGCCCTA 360
|||||

QY 4336 TCACTTCCCAATGAGGCTAAAGTTGTTTCTACTGATCAATGCCCTGACAGTTGCA 4395
|||||
Db 361 TCACTTCCCAATGAGGCTAAAGTTGTTTCTACTGATCAATGCCCTGACAGTTGCA 420
|||||

QY 4396 TTTATTTAATGAAAAAAGAACTGGGATTAATCTCTAATCAGTGAATGACATGAG 4455
|||||
Db 421 TTTATTTAATGAAAAAAGAACTGGGATTAATCTCTAATCAGTGAATGACATGAG 480
|||||

QY 4456 ACCAATGTGCTCACTTACCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 4515
|||||

Db 481 ACCAATGTGCTCACTTACCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTTC- 538
|||||

QY 4516 TAATGTGAGACAGATCTCATTTCTTTGCTTAGCTGAGTGCATGCGCAATCTCGC 4575
|||||

Db 539 TAATGTGAGACAGATCTCATTTCTTTGCTTAGCTGAGTGCATGCGCAATCTCGC 598
|||||

QY 4576 TCACTGCAACTGTGCTGCTGGGCTCAGCAATTTCCACCTCAGCTCCCAATAGC 4635
|||||
Db 599 TCACTGCAACTGTGCTGCTGGGCTCAGCAATTTCCACCTCAGCTCCCAATAGC 658
|||||

QY 4636 TGGGATCTGCGCAAAACCAACCACTAATTTGATTTTGTAGAGACAGG 4695
|||||

Db 659 TGGGATCTGCGCAAAACCAACCACTGCGCAAGTGTGATTTGTGCAAGACCGG 714
|||||

QY 4696 TTTACATGTTGGCCAGGCTGTCTCAACCTCTGGGCTCAGCAATCTC 4747
|||||
Db 715 GTTCCCATGTTTGGCCAGGTTGTCTCAACTCTCGGCTCAGCAATCTTC 766
|||||

RESULT 12
BQ950224 977 bp mRNA linear EST 21-AUG-2002
LOCUS BQ950224
DEFINITION AGENCOURT_8801117 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6311019
5', mRNA sequence.
ACCESSION BQ950224
VERSION BQ950224.1 GI:22365702
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 977)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov
Tissue Procurement: Susan L. Sullivan, PhD.
CDNA Library Preparation: Resgen, Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov
Plate: LLM13734 row: e column: 04
High quality sequence start: 7
High quality sequence stop: 496.

FEATURES
source
1. 977
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:6311019"
/clone_1lb="NIH_MGC_129"
/lab_host="DH10B (phage-resistant)"
/note="Organ: olfactory epithelium; Vector: pCMV-SPORT6.1.cdb; Site_1: EcoRV; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2.2 kb. Constructed by Resgen, Invitrogen Corp. Note: this is a NIH_MGC Library."

BASE COUNT 242 a 269 c 190 g 274 t 2 others

ORIGIN

Query Match 11.2%; Score 688.8; DB 14; Length 977;
Best Local Similarity 95.1%; Pred. No. 6e-86;
Matches 742; Conservative 0; Mismatches 34; Indels 4; Gaps 3;

QY 4214 AGATCTGCCATGCTGGGCTCACAATTTCTACCTAGCAACAACCTGGCTGAGCTGGGAC 4273
|||||
Db 8 ATNNCTGGCATGCTGGGCTCACAATTTCTACCTAGCAACAACCTGGCTGAGCTGGGAC 67
|||||

QY 4274 CAGCTGCGCTTTAGAGGGGTGCTCACTTCACTGACACAGGCTACCAAGCCCACTAGCGCC 4333
|||||
Db 68 CAGCTGCGCTTTAGAGGGGTGCTCACTTCACTGACACAGGCTACCAAGCCCACTAGCGCC 127
|||||

Accession	Source	Organism	Reference	Authors	Title	Journal	Comment
4334	TATCACTTCCCA	CAATGAGGCTAA	GTGTTGTTGTTCT	ACTGATGCAT	TGCCCGCAGGTTG	4393	
Db	128	TATCACTTCCCA	CAATGAGGCTAA	GTGTTGTTGTTCT	ACTGATGCAT	TGCCCGCAGGTTG	187
4394	CATTATTTGTA	TATGAAAAAGAA	AGAGTGGGATTT	ATATCTATATC	TATAGAGTATG	4453	
Db	188	CATTATTTGTA	TATGAAAAAGAA	AGAGTGGGATTT	ATATCTATATC	TATAGAGTATG	247
4454	AGACCAATGTG	TGCTCACATTA	CCCTTTTCTTTT	CTTTTCTTTT	CTTTTCTTTT	4513	
Db	248	AGACCAATGTG	TGCTCACATTA	CCCTTTTCTTTT	CTTTTCTTTT	CTTTTCTTTT	307
4514	TTTATGTGAG	CAGGATCTAT	CTGTGCTAGG	CGTGGAGTGC	ATGAGGCGCAT	CTCG	4573
Db	308	TTTATGTGAG	CAGGATCTAT	CTGTGCTAGG	CGTGGAGTGC	ATGAGGCGCAT	367
4574	GCTCACTGCA	ACCTTGCTGCT	CTGGGCTCA	AGCAATTTCC	CACTGACCTCC	CAATA	4633
Db	368	GCTCACTGCA	ACCTTGCTGCT	CTGGGCTCA	AGCAATTTCC	CACTGACCTCC	CAATA
4634	GCTGGATCTG	CGACAAACCA	CCATGCCAGCT	TAATTTTGTAT	TTTGTAGACAG	4693	
Db	428	GCTGGATCTG	CGACAAACCA	CCATGCCAGCT	TAATTTTGTAT	TTTGTAGACAG	487
4694	GTTTACACAT	GTTCGCCAGG	TGCTCACA	CTCTGGGCTCA	AGCAATCTCT	GCT	4753
Db	488	GTTTACACAT	GTTCGCCAGG	TGCTCACA	CTCTGGGCTCA	AGCAATCTCT	GCT
4754	CGGCTTCCCA	AAAGTCTGGG	ATTAGATGTG	AGACCGCAT	CCGACCCCA	CACCTTCA	4813
Db	548	CGGCTTCCCA	AAAGTCTGGG	ATTAGATGTG	AGACCGCAT	CCGACCCCA	CACCTTCA
4814	TTTATA	-CCAATTAC	CTGCCAGTAA	-CTGTGAC	TTTTTGCTTCTCA	-CCCTGCTCT	4869
Db	608	TTTATA	-CCAATTAC	CTGCCAGTAA	-CTGTGAC	TTTTTGCTTCTCA	-CCCTGCTCT
4870	GATCTGGA	AGAGAGGAT	TATGTTAT	AGCTTGTG	CAGACAGTCC	CAAGTTCAAT	ATTTC
Db	668	GATCTGGA	AGAGAGGAT	TATGTTAT	AGCTTGTG	CAGACAGTCC	CAAGTTCAAT
4930	TGCGGCAAAA	CTTCTCTCA	AAAAAATA	ATATGACT	TAATGATTC	CAATGCAAT	CACTT
Db	728	TGCGGCAAAA	CTTCTCTCA	AAAAAATA	ATATGACT	TAATGATTC	CAATGCAAT
RESULT 13	B0018619/c						
LOCUS	B0018619	712 bp	mrna	linear	EST 27-MAR-2002		
DEFINITION	UI-H-DHI-awu-1-06-0-UI.s1	NCI-CCAP-DHI	Homo sapiens	CDNA clone	IMAGE:5823821 3', mRNA sequence.		
ACCESSION	B0018619						
VERSION	B0018619.1	GI:19753896					
KEYWORDS	EST.						
SOURCE	human.						
ORGANISM	Homo sapiens						
REFERENCE	Mumaiyoti, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
AUTHORS	Enkhalita; Eutheria; Primates; Carnivora; Homiidae; Homo.						
TITLE	1 (bases 1 to 712)						
JOURNAL	NCI-CCAP	http://www.ncbi.nlm.nih.gov/ncicgap.					
COMMENT	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),						
	Tumor Gene Index						
	Unpublished (1997)						
	Contact: Robert Strausberg, Ph.D.						
	Email: cgapbs-remail.nih.gov						
	Tissue Procurement: Dr. Jose Mercuende						
	CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa						
	CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa						
	DNA Sequencing by: Dr. M. Bento Soares, University of Iowa						
	Clone Distribution: Clone distribution information can be found						
	through the I.M.A.G.E. Consortium/LINL at: http://image.lnln.gov						
	Seq primer: M13 FORWARD						
	POLYA-yes.						

FEATURES	source	Location/Qualifiers
		1..712
		/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/clone="IMAGE:5823821"
		/clone_lib="NCI_CGAP_DHI"
		/tissue_type="Metastatic Chondrosarcoma"
		/dev_stage="Adult"
		/lab_host="DHI0B (Life Technologies)"
		/note="Organ: Lung; Vector: pRTF3-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; NCI_CGAP_DHI is a normalized cDNA library containing the following tissue(s): VS-8 Cell line from Metastatic Chondrosarcoma in Lung. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pRTF3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is AGATCATTTGC.
		TAG_LIB=OI-H-DHI
		TAG_TISSUE=Lung
		TAG_SEQ=AGATCATTTGC"
BASE COUNT	174 a	163 c 192 g 181 t 2 others
ORIGIN		
Query Match	11.1%;	Score 687; DB 14; Length 712;
Best Local Similarity	99.0%;	Pred. No. 1.3e-85;
Matches 690; Conservative	0; Mismatches 7; Indels 0; Gaps 0;	
QY	5226	GCTTCCTCCCTCATCTGAGAGACCCCTATTTAATTCACTGTGTGAGAGCCCTAGCCCTCC
DB	699	GCTTCCTCCCTCATCTGAGAGACCCCTATTTAATTCACTGTGTGAGAGCCCTAGCCCTCC
QY	5286	ATTCTCGACATTTCCCAACCTCCAGAGCCCTTCGAAGAGAGTAGTGCCTCGCATTC
DB	639	ATTCTNGACATTTCCCAACCTCCAGAGCCCTTCGAAGAGAGTAGTGCCTCGCATTC
QY	5346	ACCAGAAGTGGAGTTGGCCCTCTTAGGTGGCTACTTTGTCACATCAACGACATCACTG
DB	579	ACCAGAAGTGGAGTTGGCCCTCTTAGGTGGCTACTTTGTCACATCAACGACATCACTG
QY	5406	TTCCTCGCAGAGACACACAGTGGCCATTTTCTTCAACTGAGAGGCTCAAAATCTCTGAC
DB	519	TTCCTCGCAGAGACACACAGTGGCCATTTTCTTCAACTGAGAGGCTCAAAATCTCTGAC
QY	5466	AAGTTGCTGTGCTCGAGAGACAGTATTTCTGTGAGGTGGCTCAAGTGAAGAGGCCACG
DB	459	AAGTTGCTGTGCTCGAGAGACAGTATTTCTGTGAGGTGGCTCAAGTGAAGAGGCCACG
QY	5526	CTGAGGAACCTCTGGCTCTTTTAAAGCCAGGCCACCTTACATAAACAATTTCAAG
DB	399	CTGAGGAACCTCTGGCTCTTTTAAAGCCAGGCCACCTTACATAAACAATTTCAAG
QY	5586	GTCACTGGAAGAGTGAAGTGCCTATTGTTGAAGCTACTGCATGCGACCCACACTGCTA
DB	339	GTCACTGGAAGAGTGAAGTGCCTATTGTTGAAGCTACTGCATGCGACCCACACTGCTA
QY	5646	TCCAGTGTGCTGCTACGACAGAGAGGCGAGGCGCATGCAAGACTGCTCTTAATGC
DB	279	TCCAGTGTGCTGCTACGACAGAGAGGCGAGGCGCATGCAAGACTGCTCTTAATGC
QY	5706	TGTGTCATTTGACAGAAAGGAGTCTCAAGAGAGACTCAACTGGACAAGCAAGCAAGC
DB	219	TGTGTCATTTGACAGAAAGGAGTCTCAAGAGAGACTCAACTGGACAAGCAAGCAAGC
QY	5766	CCACGGAGACATGCGCTTGGTAAAGTTTACACACATGTGTGTGATTTGCAAGTGTTC
DB	159	CCACGGAGACATGCGCTTGGTAAAGTTTACACACATGTGTGTGATTTGCAAGTGTTC

OY	5826	ACTGGAATTAATTATTTTCAATTCAGAACTCTTTAAGTGCGCATTTTATCATTTCCGTGG	5885
Dd	99	ACTGGAATTAATTATTTTCAATTCAGAACTCTTTAAGTGCGCATTTTATCATTTCCGTGG	40
OY	5886	CTTTAAATAAACAAATGTACCAAAAAACAAGTATCAA	5922
Dd	39	CTTTAAATAAACAAATGTACCAAAAAACAAGTATCAA	3
RESULT 14			
Bg118850		Bg118850	835 bp mRNA linear EST 30-JAN-2001
LOCUS		602347846p1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4443010 5'	
DEFINITION		mRNA sequence.	
ACCESSION		Bg118850	
VERSION		Bg118850.1 GI:12612356	
KEYWORDS		EST.	
SOURCE		human.	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS		Mammalia; Eutheria; Primates; Carnivora; Hominoidea; Homo.	
TITLE		1 (bases 1 to 835)	
JOURNAL		NIH-MGC http://mhc.nci.nih.gov/.	
COMMENT		National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgabbs@mail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov Plate: LHAM10216 row: k column: 11 High quality sequence stop: 717. Location/Qualifiers 1 . 835 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:4443010" /clone_1lb="NIH-MGC_90" /tissue_type="adenocarcinoma, cell_line" /lab_host="DH10B (phage-resistant)" /note="Organ: liver; Vector: pCMV-Sport6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.7 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH-MGC library."	
FEATURES			
source			
BASE COUNT		182 a 236 c 209 g 207 t	1 others
ORIGIN			
Query Match		11.1%; Score 683.6; DB 12; Length 835;	
Best Local Similarity		95.0%; Pred. No. 3.4e+85;	
Matches 783; Conservative		0; Mismatches 29; Indels 12; Gaps	7;
OY	5007	ACTGTTCACATGCGATTAAGAAGGATTTTATAGTCTCCTAAATGGCGTACTGCA	5066
Dd	1	ACTGTTCACATGCGATTAAGAAGGATTTTATAGTCTCCTAAATGGCGTACTGCA	60
OY	5067	AGACCTCTTGAACACTTTCAGAGAGATAGGATTTTAAGTATGCCCTTGGCGTGCCTA	5126
Dd	61	AGACCTCTTGAACACTTTCAGAGAGATAGGATTTTAAGTATGCCCTTGGCGTGCCTA	120
OY	5127	TGGCACCTTTCCTCTGAAGTCTGGTTCCTGCCAGACCTTGGCGTGTGAGCG	5186
Dd	121	TGGCACCTTTCCTCTGAAGTCTGGTTCCTGCCAGACCTTGGCGTGTGAGCG	180
OY	5187	AGATGCTACCCCTGCATTAAGGGCCAAGAGAGGCGTCCGCTTCCCTCAGTAGAGA	5246
Dd	181	AGATGCTACCCCTGCATTAAGGGCCAAGAGAGGCGTCCGCTTCCCTCAGTAGAGA	240
OY	5247	GCCCTATTGTAATTCACGTGTGGAGCCCAAGCCCTCATTCTGACATTCGCCAACCT	5306

[illegible]

/tissue_type="epithelioid carcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pORF7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACAG(C). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH-MGC Library. |"

BASE COUNT 170 a 205 c 171 g 137 t
ORIGIN

Query Match 11.0%: Score 678.8: DB 13: Length 683:
Best Local Similarity 99.7%: Pred.No.1.8e-84;
Matches 680: Conservative 0: Mismatches 2: Indels 0: Gaps 0:
QY 1103 GTGAGAGAGAGAGAGCGGTTGAATACATCCGGGCTCCACACCAACCCGAGG 1162
DB 2 GTGAGAGAGAGAGAGCGGTTGAATACATCCGGGCTCCACACCAACCCGAGG 61
QY 1163 TGTTCAGCTGAGAGCAAGACCTGGGACATGGGGGAACCTTCAACTCTCTCTGC 1222
DB 62 TGTTCAGCTGAGAGCAAGACCTGGGACATGGGGGAACCTTCAACTCTCTCTGC 121
QY 1223 AAGCGTGTACCAAGATGCGCAAGATCTCCGGCTGCGAGTTCCAAAGTTTGG 1282
DB 122 AAGCGTGTACCAAGATGCGCAAGATCTCCGGCTGCGAGTTCCAAAGTTTGG 181
QY 1283 TCCAAATCCACAAATGAAGCAATTAATCTACGTGGTGGTGAAGTGAAGAGAG 1342
DB 182 TCCAAATCCACAAATGAAGCAATTAATCTACGTGGTGGTGAAGTGAAGAGAG 241
QY 1343 CCATGTCACTCACTCGACGACGCGCCGTCAAAACAGACCGCAAGTTGTCCCTGCT 1402
DB 242 CCATGTCACTCACTCGACGACGCGCCGTCAAAACAGACCGCAAGTTGTCCCTGCT 301
QY 1403 GTTTCGTGTGTAGAAATTCGGACCTGCAAGTACCACTCACTGACATCTGGCTCA 1462
DB 302 GTTTCGTGTGTAGAAATTCGGACCTGCAAGTACCACTCACTGACATCTGGCTCA 361
QY 1463 AACACAAAATCTCTCTCTTGTGATGATCTGACACGCTGTGGATGATGGAAAAA 1522
DB 362 AACACAAAATCTCTCTCTTGTGATGATCTGACACGCTGTGGATGATGGAAAAA 421
QY 1523 CCATAAGCTGCACAGACACCGGTACTGCAAAAGAAATCTACTACTCTCAAGTGCCA 1582
DB 422 CCATAAGCTGCACAGACACCGGTACTGCAAAAGAAATCTACTACTCTCAAGTGCCA 481
QY 1583 GTGACATCTCTCACTGCTGTGGAGCTGCATGACTTCTCTGGAAGCTGTGGTCCCA 1642
DB 482 GTGACATCTCTCACTGCTGTGGAGCTGCATGACTTCTCTGGAAGCTGTGGTCCCA 541
QY 1643 AAGACAGGCTCAGCTGTGTGTGTCAGCCAGAAAGCTGCAGCAGCATACACAGAGA 1702
DB 542 AAGACAGGCTCAGCTGTGTGTGTCAGCCAGAAAGCTGCAGCAGCATACACAGAGA 601
QY 1703 AGCCCTGCACACAGCTTCACTACCTGTGGCCAGTGCCATACCCAGACGACTGT 1762
DB 602 AGCCCTGCACACAGCTTCACTACCTGTGGCCAGTGCCATACCCAGACGACTGT 661
QY 1763 ACTTGGCTCTCTTGTGCGCGG 1784
DB 662 ACTTGGCTCTCTTGTGCGCGG 683

Search completed: November 11, 2002, 11:24:42
Job time : 7744 secs

This Page Blank (uspto)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 11, 2002, 11:32:51 : Search time 77 Seconds

(without alignments)
2237.083 Million cell updates/sec

Title: US-09-899-569A-4

Perfect score: 4394

Sequence: 1 MAGLNCGVSTALLGVLLGA.....SSKDPIDPLNTQSPMEPAE 836

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTRMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_protent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeal:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4394	100.0	836	4	Q96Q07
2	4385	99.8	836	4	Q9H5V8
3	3411	77.6	649	4	Q9H8C2
4	2047	46.6	392	4	Q9H676
5	1779	40.5	343	4	Q8WU91
6	1594.5	36.3	384	11	Q921M9
7	140	3.2	746	5	001654
8	137.5	3.1	719	13	Q9PVY2
9	137.5	3.1	3623	11	070244
10	127	2.9	832	11	Q9J1R4
11	126	2.9	595	4	Q96P01
12	123.5	2.8	986	5	076335
13	123.5	2.8	3620	6	Q9T053
14	123	2.8	724	10	Q93VB7
15	118.5	2.7	1406	4	Q9Y216
16	118	2.7	2766	11	Q9QZB8

17	118	2.7	3564	11	Q923L3	Q923L3 mus musculus
18	118	2.7	3623	4	Q60494	Q60494 homo sapien
19	117.5	2.7	774	5	Q9YAA2	Q9YAA2 drosophila
20	117.5	2.7	1464	5	Q24132	Q24132 drosophila
21	117.5	2.7	1464	5	Q23995	Q23995 drosophila
22	117.5	2.7	1464	5	Q9VC47	Q9VC47 drosophila
23	117	2.7	597	6	Q9SLH0	Q9SLH0 pan troglod
24	117	2.7	603	5	Q22075	Q22075 caenorhabd
25	117	2.7	1809	5	Q9ALB8	Q9ALB8 caenorhabd
26	116.5	2.7	1027	5	Q17657	Q17657 caenorhabd
27	116.5	2.7	1424	10	Q9L7R5	Q9L7R5 arbidops
28	116.5	2.7	3121	13	Q42269	Q42269 brachydant
29	116.5	2.7	4215	5	Q9W332	Q9W332 drosophila
30	115.5	2.6	407	11	Q9D224	Q9D224 mus musculu
31	115.5	2.6	1210	5	Q8T1W5	Q8T1W5 dictyostell
32	115.5	2.6	1296	11	Q91VX3	Q91VX3 mus musculu
33	114.5	2.6	701	13	Q90X05	Q90X05 xenopus lae
34	114.5	2.6	1196	13	Q98TF1	Q98TF1 cyprinus ca
35	114.5	2.6	1573	10	Q94LS7	Q94LS7 oryza sativ
36	114.5	2.6	3898	12	Q11994	Q11994 muscosal dis
37	114	2.6	449	5	Q95RA3	Q95RA3 drosophila
38	114	2.6	16215	5	Q9NFS3	Q9NFS3 drosophila
39	113.5	2.6	688	13	Q9PVY4	Q9PVY4 xenopus lae
40	113.5	2.6	838	5	Q9VOA9	Q9VOA9 drosophila
41	113	2.6	1224	3	Q13637	Q13637 schizosacch
42	113	2.6	3139	11	Q91ZHO	Q91ZHO mus musculu
43	113	2.6	3129	11	Q8VHQ0	Q8VHQ0 mus musculu
44	112.5	2.6	701	13	Q8UVQ4	Q8UVQ4 xenopus lae
45	112.5	2.6	1018	10	Q9LSX6	Q9LSX6 arbidops

ALIGNMENTS

RESULT 1

ID	Q96Q07	PRELIMINARY:	PRT:	836 AA.
AC	Q96Q07	01-DEC-2001 (TREMBLrel. 19, Created)		
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)			
DE	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DR	CUB domain containing protein 1.			
GN	CDCP1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=21359660; PubMed=11466621;			
RA	Scherl-Mostaguer M., Sommergruber W., Abseher R., Hauptmann R.,			
RA	Ambros P., Schweifer N.;			
RT	*Identification of a novel gene, CDCP1, overexpressed in human			
RT	colorectal cancer.;			
RL	Oncogene 20:4402-4408(2001).			
DR	EMBL: AY026461; AAK02058.1; .			
SQ	SEQUENCE 836 AA; 92873 MW; FB4D2BBDD35C519 CRC64;			

Query Match	100.0%; Score 4394; DB 4; Length 836;
Best local Similarity	100.0%; Pred. No. 0;
Matches	836; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 MAGLNCGVSTALLGVLLGAARLPKGAEPFETALPRESNITVLIKIGTPTLLAKPCYIV 60
DB	1 MAGLNCGVSTALLGVLLGAARLPKGAEPFETALPRESNITVLIKIGTPTLLAKPCYIV 60
QY	61 SRKHTMTSIKSGERTVFQSCSPENHVFETQKIDMSGCPGEGVQLDPSTSLPT 120
DB	61 SRKHTMTSIKSGERTVFQSCSPENHVFETQKIDMSGCPGEGVQLDPSTSLPT 120
QY	121 LNRFTFMDVKAHKSIGLEQFSIPRLQIGPGSCPDGVTSHSGRIDATVVRIGFCSN 180
DB	121 LNRFTFMDVKAHKSIGLEQFSIPRLQIGPGSCPDGVTSHSGRIDATVVRIGFCSN 180

```

QY 181 GTVSRIKMOGVKMLHLPWFHPRNVSFGFSIANRSSIKRLCIIESYFEGSGATLMSANY 240
    |||
Db 181 GTVSRIKMOGVKMLHLPWFHPRNVSFGFSIANRSSIKRLCIIESYFEGSGATLMSANY 240
QY 241 PEQFPDELTMOFVVPVPAHLRASVSFLNFMNSCERKEEVEYYIGSTINPEVFKLEK 300
    |||
Db 241 PEQFPDELTMOFVVPVPAHLRASVSFLNFMNSCERKEEVEYYIGSTINPEVFKLEK 300
QY 301 QPNNMAGNFNLISQGCCDQDQASPGILRLQFQVLYQHPQNSNKIYYVDLSNERAMSLTIE 360
    |||
Db 301 QPNNMAGNFNLISQGCCDQDQASPGILRLQFQVLYQHPQNSNKIYYVDLSNERAMSLTIE 360
QY 361 PRPVKOSRKRVPCGFCVLESRTCSNLTTLTSSGSKHKISFLCDDLTRLMMNVERTISCTDH 420
    |||
Db 361 PRPVKOSRKRVPCGFCVLESRTCSNLTTLTSSGSKHKISFLCDDLTRLMMNVERTISCTDH 420
QY 421 RYCORSYSIQVPSDILHLPELHDEFSWKLLVPRKDRSLVLVPAOKLQOQTHHEKPCNTSF 480
    |||
Db 421 RYCORSYSIQVPSDILHLPELHDEFSWKLLVPRKDRSLVLVPAOKLQOQTHHEKPCNTSF 480
QY 481 SYLVASAIIPSDILYFSGFCGSGSIKQIQVKNISVTLTRTPAPSFQDASRQGLTVSFIPY 540
    |||
Db 481 SYLVASAIIPSDILYFSGFCGSGSIKQIQVKNISVTLTRTPAPSFQDASRQGLTVSFIPY 540
QY 541 FKEEGVFTVTPDTRKSKYLLTPTMMDRGLPSTLTVSNMISVPRQVACLTFEKKERSGVQC 600
    |||
Db 541 FKEEGVFTVTPDTRKSKYLLTPTMMDRGLPSTLTVSNMISVPRQVACLTFEKKERSGVQC 600
QY 601 TGRAFMIOQORTRAEIEFLDEDLVLPKPSFHHSFVWVNSCSPSNGKOLDLFSVTLL 660
    |||
Db 601 TGRAFMIOQORTRAEIEFLDEDLVLPKPSFHHSFVWVNSCSPSNGKOLDLFSVTLL 660
QY 661 PRTVDLTVILIAVGGVLLLSALGLIICVKKKKKTKNGPANGIYNGINTEMPPROPK 720
    |||
Db 661 PRTVDLTVILIAVGGVLLLSALGLIICVKKKKKTKNGPANGIYNGINTEMPPROPK 720
QY 721 KFOGKRNDNSHYAVIEDTMVYGHLLQDSSGSLQPEVDYTRPFQGTWGCVPSPPTIC 780
    |||
Db 721 KFOGKRNDNSHYAVIEDTMVYGHLLQDSSGSLQPEVDYTRPFQGTWGCVPSPPTIC 780
QY 781 SRAPTAKLATEEPPRPSPESESEPTFFSHPNNGDVSSKOTDIPILSTQPMPEPAE 836
    |||
Db 781 SRAPTAKLATEEPPRPSPESESEPTFFSHPNNGDVSSKOTDIPILSTQPMPEPAE 836

```

RESULT 2

```

O9H5V8 PRELIMINARY; PRT; 836 AA.
AC O9H5V8;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE CDNA: FLJ22969 fis, clone KAT10759.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Matanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA Suzuki Y., Odayashi M., Nishi T., Shibahara T., Tanaka T.,
RA Nakamura Y., Isigai T., Sugano S.;
RT "NEO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AR026622; BAB1511.1; -; 9B980475C3E5C4C8 CRC64;
SQ SEQUENCE 836 AA; 92874 MW; 9B980475C3E5C4C8 CRC64;

```

```

Query Match          99.8%; Score 4385; DB 4; Length 836;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 834; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAGLNCGVSTALLGVLLGAAARLPRAEAFAEIALPRESNITVLIKGTPTLLAKPCYIYI 60

```

```

Db 1 MAGLNCGVSTALLGVLLGAAARLPRAEAFAEIALPRESNITVLIKGTPTLLAKPCYIYI 60
    |||
QY 61 SKRHITMLSIKSGERIVTFPSCSPENHFIETLOKNIDCSGCPGCEVOLQPSSTLLPT 120
    |||
Db 61 SKRHITMLSIKSGERIVTFPSCSPENHFIETLOKNIDCSGCPGCEVOLQPSSTLLPT 120
QY 121 LNRFITWDVKAHKSIGLELOFSLPRLQIGPGSCPDGVTNHSIGRIDATVVRIGTFCSN 180
    |||
Db 121 LNRFITWDVKAHKSIGLELOFSLPRLQIGPGSCPDGVTNHSIGRIDATVVRIGTFCSN 180
QY 181 GTVSRIKMOGVKMLHLPWFHPRNVSFGFSIANRSSIKRLCIIESYFEGSGATLMSANY 240
    |||
Db 181 GTVSRIKMOGVKMLHLPWFHPRNVSFGFSIANRSSIKRLCIIESYFEGSGATLMSANY 240
QY 241 PEQFPDELTMOFVVPVPAHLRASVSFLNFMNSCERKEEVEYYIGSTINPEVFKLEK 300
    |||
Db 241 PEQFPDELTMOFVVPVPAHLRASVSFLNFMNSCERKEEVEYYIGSTINPEVFKLEK 300
QY 301 QPNNMAGNFNLISQGCCDQDQASPGILRLQFQVLYQHPQNSNKIYYVDLSNERAMSLTIE 360
    |||
Db 301 QPNNMAGNFNLISQGCCDQDQASPGILRLQFQVLYQHPQNSNKIYYVDLSNERAMSLTIE 360
QY 361 PRPVKOSRKRVPCGFCVLESRTCSNLTTLTSSGSKHKISFLCDDLTRLMMNVERTISCTDH 420
    |||
Db 361 PRPVKOSRKRVPCGFCVLESRTCSNLTTLTSSGSKHKISFLCDDLTRLMMNVERTISCTDH 420
QY 421 RYCORSYSIQVPSDILHLPELHDEFSWKLLVPRKDRSLVLVPAOKLQOQTHHEKPCNTSF 480
    |||
Db 421 RYCORSYSIQVPSDILHLPELHDEFSWKLLVPRKDRSLVLVPAOKLQOQTHHEKPCNTSF 480
QY 481 SYLVASAIIPSDILYFSGFCGSGSIKQIQVKNISVTLTRTPAPSFQDASRQGLTVSFIPY 540
    |||
Db 481 SYLVASAIIPSDILYFSGFCGSGSIKQIQVKNISVTLTRTPAPSFQDASRQGLTVSFIPY 540
QY 541 FKEEGVFTVTPDTRKSKYLLTPTMMDRGLPSTLTVSNMISVPRQVACLTFEKKERSGVQC 600
    |||
Db 541 FKEEGVFTVTPDTRKSKYLLTPTMMDRGLPSTLTVSNMISVPRQVACLTFEKKERSGVQC 600
QY 601 TGRAFMIOQORTRAEIEFLDEDLVLPKPSFHHSFVWVNSCSPSNGKOLDLFSVTLL 660
    |||
Db 601 TGRAFMIOQORTRAEIEFLDEDLVLPKPSFHHSFVWVNSCSPSNGKOLDLFSVTLL 660
QY 661 PRTVDLTVILIAVGGVLLLSALGLIICVKKKKKTKNGPANGIYNGINTEMPPROPK 720
    |||
Db 661 PRTVDLTVILIAVGGVLLLSALGLIICVKKKKKTKNGPANGIYNGINTEMPPROPK 720
QY 721 KFOGKRNDNSHYAVIEDTMVYGHLLQDSSGSLQPEVDYTRPFQGTWGCVPSPPTIC 780
    |||
Db 721 KFOGKRNDNSHYAVIEDTMVYGHLLQDSSGSLQPEVDYTRPFQGTWGCVPSPPTIC 780
QY 781 SRAPTAKLATEEPPRPSPESESEPTFFSHPNNGDVSSKOTDIPILSTQPMPEPAE 836
    |||
Db 781 SRAPTAKLATEEPPRPSPESESEPTFFSHPNNGDVSSKOTDIPILSTQPMPEPAE 836

```

RESULT 3

```

O9H8C2 PRELIMINARY; PRT; 649 AA.
AC O9H8C2;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE CDNA FLJ13772 fis, clone PLACE4000300.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA ISOGAI T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,

```


RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
 RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
 RA Makamatsu A., Nakamura Y., Nagahari K., Masuno Y., Oshima A.,
 RT "NEO human cDNA sequencing project."
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK023834; BAB1695.1; -
 FT NON_TER 392
 SO SEQUENCE 649 AA; 72579 MW; 0F4404AD31D60E9A CRC64;

Query Match 77.6%; Score 3411; DB 4; Length 649;
 Best Local Similarity 99.5%; Pred. No. 1.3e-293;
 Matches 646; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 188 MQGVKALHLPFHPHNSGFSIANSRSIKRCLITISVEGEGSATLMSANYPEGEPD 247
 |||||
 DB 1 MQGVKALHLPFHPHNSGFSIANSRSIKRCLITISVEGEGSATLMSANYPEGEPD 60
 |||||
 QY 248 ELMTQFVPAHLRASVSLNLSNCRKEEREVEYIPGSTNPEYFKLEDKOPGMAG 307
 |||||
 DB 61 ELMTQFVPAHLRASVSLNLSNCRKEEREVEYIPGSTNPEYFKLEDKOPGMAG 120
 |||||
 QY 308 NEFLSLGGCDQDQSPGILRLQFOVLVQHPQNSNKITYVDLSNERAMSLTIEPRPV 367
 |||||
 DB 121 NEFLSLGGCDQDQSPGILRLQFOVLVQHPQNSNKITYVDLSNERAMSLTIEPRPV 180
 |||||
 QY 368 RKEVPCFVCELESRTCSNLTLSGSKHKISFLCDDLRLMMNVEKTICTDHRVCO 427
 |||||
 DB 181 RKEVPCFVCELESRTCSNLTLSGSKHKISFLCDDLRLMMNVEKTICTDHRVCO 240
 |||||
 QY 428 YSLQVPSDILHLPVELHDFSWKLLVPRKRLSLVLPQAKLQOHTHEKPCNTSFLV 487
 |||||
 DB 241 YSLQVPSDILHLPVELHDFSWKLLVPRKRLSLVLPQAKLQOHTHEKPCNTSFLV 300
 |||||
 QY 488 IPSODILFSGFCGSGSIKOIOVKONISVTLRTFAPSQOESRQGLTVSFIPFKE 547
 |||||
 DB 301 IPSODILFSGFCGSGSIKOIOVKONISVTLRTFAPSQOESRQGLTVSFIPFKE 360
 |||||
 QY 548 TVTPDTSKYVLRTPMMDRGLPSLTSVSWNISVPRDOVACLTFFKERSGVVCOGRA 607
 |||||
 DB 361 TVTPDTSKYVLRTPMMDRGLPSLTSVSWNISVPRDOVACLTFFKERSGVVCOGRA 420
 |||||
 QY 608 IOEQRRAEIRFSLDEVDLPKPSFHHSFWNINSNCSPTSGKLDLFLSVTLPR 667
 |||||
 DB 421 IOEQRRAEIRFSLDEVDLPKPSFHHSFWNINSNCSPTSGKLDLFLSVTLPR 480
 |||||
 QY 668 VILIAVGGVLLSLGLITICCVKKKKTKNGRAVGITNGINTEMPQPKFQGRK 727
 |||||
 DB 481 VILIAVGGVLLSLGLITICCVKKKKTKNGRAVGITNGINTEMPQPKFQGRK 540
 |||||
 QY 728 DNDSHYAVIEDTMYGHLQDSSGFLQPEVDYRPFQGTMGVCPPTICSRAPTAK 787
 |||||
 DB 541 DNDSHYAVIEDTMYGHLQDSSGFLQPEVDYRPFQGTMGVCPPTICSRAPTAK 600
 |||||
 QY 788 LATEPPPSPESESEPTFSHPNNGDVSSKDTDIPILNTQPMPEAE 836
 |||||
 DB 601 LATEPPPSPESESEPTFSHPNNGDVSSKDTDIPILNTQPMPEAE 649
 |||||

RESULT 4

Q9H676 PRELIMINARY; PRT; 392 AA.

AC Q9H676
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE CDNA: FLJ22534 f1s, clone HRC13020 (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kawabata A., Hiki T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
 RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,

RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
 RT "NEO human cDNA sequencing project."
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK026187; BAB1538.1; -
 FT NON_TER 392
 SO SEQUENCE 392 AA; 44031 MW; 3E21C8B2703E52F3 CRC64;

Query Match 46.6%; Score 2047; DB 4; Length 392;
 Best Local Similarity 99.7%; Pred. No. 6.4e-173;
 Matches 391; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 305 MAGFNLSLGGCDQDQSPGILRLQFOVLVQHPQNSNKITYVDLSNERAMSLTIEPRPV 364
 |||||
 DB 1 MAGFNLSLGGCDQDQSPGILRLQFOVLVQHPQNSNKITYVDLSNERAMSLTIEPRPV 60
 |||||
 QY 365 KQSRKFFVPCFVCELESRTCSNLTLSGSKHKISFLCDDLRLMMNVEKTICTDHRVCO 424
 |||||
 DB 61 KQSRKFFVPCFVCELESRTCSNLTLSGSKHKISFLCDDLRLMMNVEKTICTDHRVCO 120
 |||||
 QY 425 RKSYSLOVPSDILHLPVELHDFSWKLLVPRKRLSLVLPQAKLQOHTHEKPCNTSFLV 484
 |||||
 DB 121 RKSYSLOVPSDILHLPVELHDFSWKLLVPRKRLSLVLPQAKLQOHTHEKPCNTSFLV 180
 |||||
 QY 485 ASAIPSDILFSGFCGSGSIKOIOVKONISVTLRTFAPSQOESRQGLTVSFIPFKEE 544
 |||||
 DB 181 ASAIPSDILFSGFCGSGSIKOIOVKONISVTLRTFAPSQOESRQGLTVSFIPFKEE 240
 |||||
 QY 545 GVEFTVPDTSKYVLRTPMMDRGLPSLTSVSWNISVPRDOVACLTFFKERSGVVCOGRA 604
 |||||
 DB 241 GVEFTVPDTSKYVLRTPMMDRGLPSLTSVSWNISVPRDOVACLTFFKERSGVVCOGRA 300
 |||||
 QY 605 FMIIQQRRAEIRFSLDEVDLPKPSFHHSFWNINSNCSPTSGKLDLFLSVTLPR 664
 |||||
 DB 301 FMIIQQRRAEIRFSLDEVDLPKPSFHHSFWNINSNCSPTSGKLDLFLSVTLPR 360
 |||||
 QY 665 DLTVILIAVGGVLLSLGLITICCVKKKK 696
 |||||
 DB 361 DLTVILIAVGGVLLSLGLITICCVKKKK 392
 |||||

RESULT 5

Q8WU91 PRELIMINARY; PRT; 343 AA.

AC Q8WU91
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Similar to hypothetical protein FLJ22969.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Tissue-Kidney;
 RA Strausberg R.;
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC021099; AAH21099.1; -
 KW Hypothetical protein.
 SO SEQUENCE 343 AA; 37817 MW; 3E4E13379DD94D1B CRC64;

Query Match 40.5%; Score 1779; DB 4; Length 343;
 Best Local Similarity 99.4%; Pred. No. 3.1e-149;
 Matches 341; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGLNCVSTALGVLIGARLPGRGAPEILAPRESNTVLIKIGTPLLAKPCYIV 60
 |||||
 DB 1 MAGLNCVSTALGVLIGARLPGRGAPEILAPRESNTVLIKIGTPLLAKPCYIV 60
 |||||
 QY 61 SKRHITMLSIKSGERIVTFSCSPENHFVIEIOKNIDCSGCPGEGVQLOPSTSLP 60
 |||||
 DB 61 SKRHITMLSIKSGERIVTFSCSPENHFVIEIOKNIDCSGCPGEGVQLOPSTSLP 60
 |||||

```

QY 121 LNRFTMDVAKHKSIGLELOPISRLRQIGGSCPDGVTHTSIGRIDATVVRIGTCRSN 180
DB 121 LNRFTMDVAKHKSIGLELOPISRLRQIGGSCPDGVTHTSIGRIDATVVRIGTCRSN 180
QY 181 GTVARIKMOEYKALHLPMPHPNVSQFSIANKSSIKRLCIISVFEQSGATILMSANY 240
DB 181 GTVARIKMOEYKALHLPMPHPNVSQFSIANKSSIKRLCIISVFEQSGATILMSANY 240
QY 241 PEGPEDELMTQFVPAHLRAVSFLNLSNCRKEERYEYIIPGSTNPEVEKLEDK 300
DB 241 PEGPEDELMTQFVPAHLRAVSFLNLSNCRKEERYEYIIPGSTNPEVEKLEDK 300
QY 301 QPGMAGNFNLISLOGCDODASQFILRQFOYLYVQHPONESNK 343
DB 301 QPGMAGNFNLISLOGCDODASQFILRQFOYLYVQHPONESNK 343
RESULT 6
QY 0921M9 PRELIMINARY; PRT; 384 AA.
AC 0921M9;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Unknown (Protein for IMAGE:370937) (Fragment).
GN AA409659.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC01340; AA41340.1;
DR MGD: MGI:2142933; AA409659.
FT NON_TER 1
SQ SEQUENCE 384 AA; 42406 MW; 4B5F17D30FE16080 CRC64;

Query Match 36.3%; Score 1594.5; DB 11; Length 384;
Best Local Similarity 77.8%; Pred. No. 8.6e-133;
Matches 301; Conservative 33; Mismatches 48; Indels 3; Gaps 3;

QY 450 LLVPRKRLSLVPAQKIQHTHEKPCNTSFSYLVASAIQSDLYFGSFCGSGSIKQIQV 509
DB 1 LLVPRKRLSLVPAQKIQHTHEKPCNTSFSYLVASAIQSDLYFGSFCGSGSIKQIQV 509
QY 510 KQNTSVTLRFPAPSGOQASROGLTVSFIPEKKEGVFTVTPDKSKVYLPTPMMDGLP 569
DB 61 KQNTSVTLRFPAPSGOQASROGLTVSFIPEKKEGVFTVTPDKSKVYLPTPMMDGLP 569
QY 570 SLTSVSNISVPRDQVACITFEKERSGVQCTGRAFMIIQORTRAEIEFSLDEVDLPRK 629
DB 121 ALSSVSNISVPRDQVACITFEKERSGVQCTGRAFMIIQORTRAEIEFSLDEVDLPRK 629
QY 630 SFHHSHFWNISNCSPTSGKQDLLEFVTLTPRTVDTLLIAVGGVLLLSALGLTIC 689
DB 181 SFHHSHFWNISNCSPTSGKQDLLEFVTLTPRTVDTLLIAVGGVLLLSALGLTIC 689
QY 690 CVKRRKKTNGKAVGIYNGINTEMERQPKKQKGRKDNDSHYAYIEDIMVGHLLD 749
DB 240 FV-KRRKKTNGKAVGIYNGINTEMERQPKKQKGRKDNDSHYAYIEDIMVGHLLD 749
QY 750 SSGSFLQPEVDVYRPGQTMGVCPPSPPTICSRAPTKLALEPPPPSPSESEPTTFS 809
DB 298 SSGSFLQPEVDVYRPGQTMGVCPPSPPTICSRAPTKLALEPPPPSPSESEPTTFS 809
QY 810 HPNNGDVSSKDTDIPLLNTQEPMPAE 836
DB 358 HPNNGDVSSKDTDIPLLNTQEPMPAE 836
RESULT 7

```

```

001654
AC 001654; PRELIMINARY; PRT; 746 AA.
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE AsnA8Pa.
OS Halocynthia roretzi (Sea squirt).
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea,
OC Stolidobranchia; Pyrosidae; Halocynthia.
OX NCBI_TaxId=7729;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HEPATOPANCREAS;
RX MEDLINE=9732374; PubMed=9177219;
RA Uchi X, Azumi K., Sasaki M., Nonaka M.;
RT "Ancient origin of the complement lectin pathway revealed by molecular
RT cloning of mannan binding protein-associated serine protease from a
RT urochordata, the Japanese Ascidian, Halocynthia roretzi."
RL Proc. Natl. Acad. Sci. U.S.A. 94:6340-6345(1997).
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.-ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL: D88204; BA19762.1; -.
DR HSP; P00763; IDPO.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR000859; CUB domain.
DR InterPro: IPR001881; EGF Ca.
DR InterPro: IPR001254; Ser_protease_Try.
DR InterPro: IPR000436; Sush1_SCR_CCP.
DR Pfam: PF00431; CUB; 2.
DR Pfam: PF00084; sush1; 1.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00032; CCP; 2.
DR SMART: SM00042; CUB; 2.
DR SMART: SM00179; EGF_CA; 1.
DR SMART: SM00020; Tryp_Spc; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; 1.
DR PROSITE: PS0180; CUB; 2.
DR PROSITE: PS00240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR EGF-like domain; Glycoprotein; Hydrolase; Hydroxylation;
KW Serine protease.
SQ SEQUENCE 746 AA; 84127 MW; 32B075317CF173E1 CRC64;

Query Match 3.28%; Score 140; DB 5; Length 746;
Best Local Similarity 19.3%; Pred. No. 0.0029;
Matches 102; Conservative 73; Mismatches 178; Indels 176; Gaps 25;

QY 220 LCIEVEFEGSGATILMSA-----NYPEGPEDELMTQFVPAHLRAVSFLNPL 271
DB 11 LLYCNVVF-----SALLTAHFGNFPSPYPRSYDINSULTNINVOGYRMSIFSTFDL 66
QY 272 -----SNCRKEERYEYIIPGSTNPEVEKLEDKQPGMAGNFNLISLOGCDODASQ 323
DB 67 EDSEYEDIGSC-----YVDVETTESNKTYAK-----FCGNYQL-----PFLDAPNP 108
QY 324 GILRQOVLVQHPONESNKIYVDLS-----NERAMSLTIERRPYKQSKK----- 369
DB 109 -----SKFTYTQNEVRVTFVSDYSISLSGFOAHYAQIDINCELMETKRTIIEEDMD 161
QY 370 -----FVPCFCLEBRTCCSSNLTLSGSKHISFLCDLTLRMNVEXTISCTD 419
DB 162 ELVVCSHYCRNVPGSYVC-----GCRPKFTLDANRHTCVASFCEN-----OVLTDON 208
QY 420 HRYCQKSYSLQVSDLIHLPELHDSWKLVLPRKD-RLSLVPAQKIQHTHEKPCNT 478
DB 209 SGHISSEFP-----ELYAKLTDCSWITQLREGLSVNLIFRAGIEHDEE-----GC 257
QY 479 SFSYLVASAIQSDLYFGSFCPGG-----SIKQIQVKNISVTLTPAPSPFOQ----- 526

```

```

Db 258 YDRLRLVLEHKSTTDYICGNQAPGNGTVMPMNTNLVOLKFTHTDLSEVKKGSFVYSTRK 317
OY 527 -----EASHQGLVSRF---PYKEGCVTVTPDTSKYYLTPMNDRLPLSTYS-5 575
Db 318 CHAHLDPNNG-SLSFSHSRSYHEFEDVATFSCDRGFDT-----GVPRKCLSDS 368
OY 576 WNI5VPRDGV-AC-----LTFKERSGVVCO-----TGAFMTI 608
Db 369 WEHSAPICQIKSCGVQFLDLPNSHIVYENSKTYSVLDVTCNOMIGMISGASKWIC 428
OY 609 QEQRTAEIFSLDEVDLPKPSFHHSFVNIN-----CSPTSGKOL 651
Db 429 ENSKIMTE-----HGLGLVAINNFNKRKPCPKICKKL 460

```

RESULT 8

```

O9PVY2 PRELIMINARY: PRT: 719 AA.
ID O9PVY2
AC 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Manose-binding lectin-associated serine protease.
OS Triakis scyllium (Leopard shark) (Triakis scyllia).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Galeomorphi; Galeoidea; Carcharhiniformes; Triakidae;
OC Triakis.
OX NCBI_TaxID=30494;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE-99008558; PubMed-9794427;
RA Endo Y., Takemashi M., Nakao M., Saiga H., Sekine H., Matsushita M.,
RA Nonaka M., Fujita T.;
RT "Two lineages of manose-binding lectin-associated serine protease
RT (MASP) in vertebrates.";
RL J. Immunol. 161:4924-4930(1998).
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRIPSTIN FAMILY.
DR EMBL: AB009074; BAA68667.1; -.
DR HSSP: P00736; IAPQ.
DR MEROPS: S01.132; -.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR000859; CUB_domain.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR003006; IG_MHC.
DR InterPro: IPR001254; Ser_protease_Try.
DR InterPro: IPR000436; Sushl_SCR_CCP.
DR Pfam: PF00431; CUB; 2.
DR Pfam: PF00084; Sushl; 2.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00032; CCP; 2.
DR SMART: SM00042; CUB; 2.
DR SMART: SM00179; EGF_Ca; 1.
DR SMART: SM00020; Tryp_Spc; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS01187; EGF_CA; 1.
DR PROSITE: PS00290; IG_MHC; UNKNOWN_1.
DR PROSITE: PS0240; TRYPSTIN_DOM; 1.
DR PROSITE: PS00135; TRYPSTIN_SER; UNKNOWN_1.
KW Calcium-binding; EGF-like domain; Glycoprotein; Hydrolase; Lectin;
KW Protease; Repeat; Serine protease.
SQ SEQUENCE 719 AA: 80425 MW: 1766415187044455 CRC64;

```

Query Match 3.1%; Score 137.5; DB 13; Length 719;
 Best Local Similarity 19.5%; Pred. No. 0.0045;
 Matches 100; Conservative 62; Mismatches 149; Indels 201; Gaps 27;

```

OY 235 IMSANYPEGEPEDELMTQFVVPAPHLRASVFLNFINLN-----CE-----RKEERYEX 284
Db 29 IQSPNTPPEGIPSDLVNTWNTTYPAGRIKILYFHHFDLEPYLEYVYKVSDELLTF 88
OY 285 IPGSTNPEYFKLEDKQPGMAGNFNLSLOGCDQDQASPGILRLQVLYVHPONSNKI 344
Db 89 CGRESTDTE-----QIRG-----RE 103
OY 345 YVVDLSNERAMSLTTPRPYKQSRKFPVCGFVLESRTCSNLTLSGSKHKSIFLCDL 404
Db 104 EV--LINSNTMAVTLKTDFSNEER-----FTGEFAHYSAYDID-BCEVKIDBALCDHF 154
OY 405 TRLMNVETISC-----TDHRYCO-----RKSYSLOVP-----SDILHL 439
Db 155 CHNYIG-GYICSCRFSEYILHTDNRTCKVECNQNVFTKSGVINSPIPNPYPSSDCLY- 212
OY 440 PVELHD-FSWKLLVPKRLSLVLPAPKLOQHTHERPKCNTSFSYLVASAIPSDLYGSP 498
Db 213 RILEEGFA-----ISLEFADMELEDHP-DVPC--PYDYL--KIKAGSMAFGPF 257
OY 499 CPBGSLKQIQVONISVTLTFTAPSFQOENSRQGLVSRF-----PYKEGCVTVTP- 551
Db 258 CGDKSPGKIDTNS-SVQILFHSG--SGDNQGMKMSFSLVGTGPIILHPANGTIEPV 313
OY 552 ---DTRSKVYLTTPMNDRLPLSTYSV-----SWNISVPRDOVA-C-----587
Db 314 QDOYSEKDOYQIKCNIGFFVKAQDEYAYQICQSGSSSIPTCKYIDCGAPPLD 373
OY 588 -----LTFKERSGVVCO-TGAFMTI0BQRTAEIFSLDEVDLPKPSFHH-634
Db 374 GITFHSKENVLTFRSHIEVS-----PALHHLER 404
OY 635 -----SWVN-----ISNCSPTSGK 649
Db 405 ELNTTYTCMENGFWVDELGTDLPTCOPVCGR 436

```

RESULT 9

```

O70244 PRELIMINARY: PRT: 3623 AA.
ID O70244
AC 070244;
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-JAN-1999 (Tremblrel. 09, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Intrinsic factor-B12 receptor precursor.
OS CUBILIN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE-98148073; PubMed-9478979;
RA Mostlunp S.K., Koziyaki R., Kristiansen M., Kayser J.H.,
RA Rasmussen H.H., Brault D., Pontillon F., Gode F.O., Christensen E.I.,
RA Hammond T.G., Verroust P.J.;
RT "The intrinsic factor-vitamin B12 receptor and target of teratogenic
RT antibodies is a megalin-binding peripheral membrane protein with
RT homology to developmental proteins.";
RL J. Biol. Chem. 273:5235-5242(1998).
CC -1- SIMILARITY: CONTAINS 26 CUB DOMAINS.
DR EMBL: AF022247; AAC71661.1; -.
DR HSSP: P00740; IEDM.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000859; CUB_domain.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR00742; EGF-2.
DR InterPro: IPR001881; EGF_Ca.
DR Pfam: PF00431; CUB; 27.
DR Pfam: PF00008; EGF; 7.
DR SMART: SM00042; CUB; 26.
DR SMART: SM00179; EGF_CA; 4.
DR SMART: SM00001; EGF_like; 4.

```

DR PROSITE; PS00010; ASX_HYDROXYL; 3.
DR PROSITE; PS01180; CUB; 27.
DR PROSITE; PS00022; EGF_1; UNKNOWN_4.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS01187; EGF_CA; 4.
KM Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation;
KW Receptor; Repeat; Signal.
FT SIGNAL; 1 20
FT CHAIN; 21 3623
SQ SEQUENCE 3623 AA; 398981 MW; 39FB792AC6545240 CRC64;

Query Match
Best Local Similarity 18.9%; Score 137.5; DB 11; Length 3623;
Matches 151; Conservative 77; Mismatches 283; Indels 289; Gaps 34;

QY 70 IKSGRIFFVFCQSPNNHVIETIKNIDCMGCPGPEVOLQPS--LLPTLNRFFI- 126
DB 2306 VSSNERIYLFHFDGSSYMGFKAKYSIACGGIVS-GDSGVIESIGYPLIPYANNVFCQ 2304
QY 127 MDVKAHKSIGLEIQFSIPRLRQIGSGSCPDGVT---HSISGRIDATVVRIGTFCNGT 182
DB 2365 WFIRGLPGHYLTLSFEDFNL-QSSPG--CTKDFEIVEINHTSGRV-----LGRYCGNST 2415
QY 183 VSRKMEGKMAHLHFWFPRNVSFGSIANKRSIKRLCTIESFEQE---GSATLMSAN 239
DB 2416 PSSVDYSSNV-ASVKFVTDSSVTASGFRLOFKRS-RQVC-----GGDLHGPTGFTSPN 2467
QY 240 YPEGFPEDELTMOQVVPVPAHLRASVSFLNPLNSL-----C 274
DB 2468 YPNPNPARKCEWITIQEGRRLVLTFTNLRLSTOPSCNSEHLIVFGINSNPLKLC 2527
QY 275 ERKEERYEYIPGSTNPEVFKLEKOP-----GNMAGNPLNSLOGCDQDAQS 322
DB 2528 SRVAVTNEFKSSGNTMK-VVEFTDGSRPYGGFTASYSTEDAVGGFLPSVSG--GNFSS 2584
QY 323 PGI-----LRLOFOVLVOHPQNESNNTIYVDLSNEBAMSLTIERRPKQSRKRVPG 373
DB 2585 PGYNGIRDYARNLDCENTLNPNNENSSISYLELISIESHODCT----- 2629
QY 374 CFVCLSESTCSSNLTLTSGSKHKISFLCDDTLRLMMNVEKTSCTDHRRCORKSYSLOVP 433
DB 2630 -FDVLEFRVGDADGPL-----IEK-----FC-----SLSAP 2654
QY 434 SDILHLP-----VELHDP----- 446
DB 2655 TAPLVIPQVWIHFVSNERYEYGFYIEYFTDCGIGRTGDNGVISSPNYPLYSAMTH 2714
QY 447 -SWKLLVPKRLSLVLPARKLOQHTHEKCNISFSYLVASAIPSOQLYFGSPGGSIK 505
DB 2715 CSWLLKAPBEHGTTLTFLSDPLEAH---PCTJSDSVTVNRGDSPGSPVIGRYCGQSVPR 2770
QY 506 QIQVKON-ISTVLTFAPSFQOEASROGLIVSFIYPKEE---GVFTVPTDKSKVYL 559
DB 2771 PIQSGSNQLVTFNT-----NNGQYRGFYATWTITNALGGGFGFHSANGF-----I 2816
QY 560 RTPNMDRGLSLISVS-----NISVRQOVACL-TFFERKSG----- 596
DB 2817 KSPHMPQTFPENSRCSTVITTHESKHWELSFDSNFRIPSSDSQCSQNSFVYVMEGRMLNK 2876
QY 597 -----VVCOTGAFMIQORRAEELFS----- 620
DB 2877 TLLATSCGDVAPSPVITVSGNIFTAVFQSEMAAGFSASFLSRGRTFNTSPGDIISPNE 2936
QY 621 -----LDED-----VLKPSRF-----HHSFWNISNCSPSGKQ 650
DB 2937 PKQYDNMNMCTYLIDAPQSLVITLTFVSHLEDSALITGCHDGLAI-IGKRLMSSTPL 2995
QY 651 LDLLFSVTLPRVLDLTVIL 670
DB 2996 VTIGSGEFLRLRYVDGPVLL 3015

RESULT 10

Q9JTB4
ID Q9JTB4 PRELIMINARY; PRT; 832 AA.
AC Q9JTB4;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Cubilin (Fragment).
GN CUBN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=ECTOPLACENTAL CONE;
RA Hammed S.M., Barth J., Knaak C., Argaves W.S.;
RT "Megalin acts in concert with cubilin to mediate endocytosis of high
RT density lipoproteins";
RL J. Biol. Chem. 0:0-0(2000).
CC EMBL; AF197159; AAF61487.1; -.
DR MGD; MGI:1931256; Cuhn.
DR InterPro; IPR000859; CUB domain.
DR Pfam; PF00431; CUB; 7.
DR SMART; SM00042; CUB; 6.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; UNKNOWN_1.
DR PROSITE; PS01180; CUB; 7.
FT NON TER 1
SQ SEQUENCE 832 AA; 92028 MW; 6B1EF8B067B47BE CRC64;

Query Match
Best Local Similarity 2.9%; Score 127; DB 11; Length 832;
Matches 94; Conservative 62; Mismatches 156; Indels 124; Gaps 22;

QY 232 SATLMSANYPGFPEDLTMOQVVPVPAHLRASVSF-LNPLN-----SNCERKEEYV----- 281
DB 22 NGTISKSPHWQTFPENSRCSTVITTHESKHWELSFDSNFRIPSSDSQCSQNSFVYVMEGRML 81
QY 282 ---EYIIPGSTN-----PEVFKLEKOPGNMAGNPLNSLOGCDQDAQSPGI 325
DB 82 ETNALLATSCGNVAPSPVITLGNIFTAVFQSEMAAGFSASF-ISRGRFTVNSGDI 140
QY 326 LRLOFOVLVOHPQNESKTIYVDLSNEBAMSLT-----IEPRYKQSRKRVPGCVLES 380
DB 141 VSPNFP---KHYDNMNMCTYLIDAPQSLVITLTFVSHLEDSALVSGTCDYDGLIIGH 197
QY 381 RTGSSNLTLSGSK-----HKSFLCDDTLRLMMNV----- 411
DB 198 NLSSTPLVTTGSELTPLRLTIDGVMNLFYSDAYITPGEFISTYRVANCGGITSIGTYGV 257
QY 412 -EKTI SCTDHR---YCORKSYSLQVPSDILHLPVELHDFSKMLVPKDRSLVLPARKL 467
DB 258 NSPFSYSTINPNNNYC---YSLQVRNDRLLT-LRFNDFE---IVPSNLCS----- 301
QY 468 QQHTHEKPCNTSFYLVASAIIP-QDLYFGSFCFGGSIKQIQVKONISVTLRTAPSFQ 526
DB 302 ---HD-----YLEVFDPGSGNRSIGKFC-GSTLIPQYIKSTNSLT-----LFT 343
QY 527 EASR--OGLVYFIP-----YKREGVFTVPTDTSKYVLRPNPMDRGLPSLSV 574
DB 344 DSSQTANGKAVSFRETIGPQCGCGYLTEDSKSFVSPDHSDGL-----YDKGLNCI--- 395
QY 575 SWNISVRQOVACLTF 590
DB 396 -WYIAPENKLVKLTFF 410

RESULT 11
Q96P01
ID Q96P01 PRELIMINARY; PRT; 595 AA.
AC Q96P01;
DT 01-DEC-2001 (Tremblrel. 19, Created)

DT 01-DEC-2001 (TReMBLrel. 19, last sequence update)
 DT 01-MAR-2002 (TReMBLrel. 20, last annotation update)
 DE Silescu-L.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21523976; PubMed=11546777;
 RA Angula T., Varki N.M., Varki A.;
 RT "A second uniquely human mutation affecting sialic acid biology";
 RL J. Biol. Chem. 276:40282-40287(2001).
 DR EMBL: AF282256; AAK71521.1;
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR003596; Ig_V.
 DR Pfam: PF00047; Ig_4.
 DR SMART: SM00406; IGV. 2.
 SQ SEQUENCE 595 AA; 64984 MW; D05662176274C5C3 CRC64;

Query Match 2.9%; Score 126; DB 4; Length 595;
 Best Local Similarity 19.9%; Pred. No. 0.035;
 Matches 135; Conservative 86; Mismatches 247; Indels 212; Gaps 34;

OY 225 SVEEGESATLMSANYPEGFEDELMTQFVPAH---LRASVFLNLSNCRKEERY 281
 DB 33 TQEGCLVSLCSFSPON-----GWTASDPVHGWTWRAG-----DHV 70
 OY 282 EYTPGSTNPEVFKLEDKOPGNMAGNFNLSLOGCDODASPGILRQFVLQHPONES 341
 DB 71 SRNIPVATNNPARAVOEE-----TDRFR-LIGDPQNKD 103
 OY 342 NKIIYVDLSNERAMSLTIERPVKOSKRVPGCFV-CLSRPSSNLTLSGKHKISFL 400
 DB 104 CTL-----SIRDRESADAGTYVFCVERGNMKMYKKYDQLSVN-VTAS 144
 OY 401 CDDLRFLMNVKRTISCTDHRQKRSYSLOVPSDLH-----LPV-----ELHDE 446
 DB 145 QDLSTKYLEVPESVTV-----QGLCVSPCSVLYPHYMTASSPYGSKFKGADI 197
 OY 447 SKMLLVKDRLSLVLPACKLOQHTH-----EKPCNTSFSYVASAIPSDLYGSE 498
 DB 198 PNDIPVATN-----TPSKVDQEDTHGRFLLDGDPQNNCSLSIRARRGDSGKYFYQ-- 249
 OY 499 CEGGSK--QIOVKONISTYLTFAFSFO-----QEASRQGLTVSIFPYFKEEG----- 545
 DB 250 VERGSRKNMYIYDKLSVHTALTALHMTPTFSIPGLSGHPRNLCS--VPMACEQGTPTIT 308
 OY 546 -----VFTVTPD-TKSKVYLTTPN-WDRGLPSLT-----SVSNMISVPRQ 584
 DB 309 WNGASVSLDPTITSSMSLSLIPQODHG-TSLTCOVTLPGAGVTMTRAVRLNISTYPPON 367
 OY 585 VACLTFEKE-RSGVVCOTGRAFMIIQORTRAEIFSLDED----- 624
 DB 368 LMTVVGQDGTASTLRNSALSYLEGQSLH--VCAYVSNPRARLSMTWMTGSLTSPSS 425
 OY 625 -----VLPKSFH---HNSWVNISNCSPTSGQDLDFSV--TLTPRYDVLVILAAVG 675
 DB 426 SNLGELEPRVHVKEGEFTCRAN--PLGSHISLSLSLQNEYTKMPSISVTLGAFG 483
 OY 676 GG---VLLSALGLIICCVKKKKKKTKNKRAVGIVNGNINTERPRPKFKOKRKNDSH 732
 DB 484 GAGATALLVLYCIFVNVVRSCKKKSAR-PAVGV-GDTGME----- 522
 OY 733 VYAVIEDTVYVGHLLQDSSGSPLOPEVDYVREFQGTMGVCPSPPTICSHAPYAKLATEE 792
 DB 523 -----DANVVRG---SASGRLIESPADDSRPHNARPALATSPRE--GEIQVASLSFKH 572
 OY 793 PPRPRPSESEPEYTFSSH 812
 DB 573 ARPOYPOEOEAIGYEYSEIN 592

RESULT 12

076355

PRELIMINARY;

PRT;

986 AA.

AC 076355;

DT 01-NOV-1998 (TReMBLrel. 08, Created)

DT 01-NOV-1998 (TReMBLrel. 08, last sequence update)

DT 01-JUN-2001 (TReMBLrel. 17, last annotation update)

DE C45G7.5 protein.

GN C45G7.5.

OC Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;

OC Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RX MEDLINE=94150718; PubMed=7906398;

RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,

RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,

RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,

RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,

RA Lightning J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,

RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkneen R.,

RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,

RA Thiermy-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,

RA Watson A., Welstock L., Wilkinson-Sproat J., Wohlman P.;

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

RT elegans.";

RL Nature 368:32-38(1994).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA Dante M., Wamsley P.;

RT "The sequence of C. elegans cosmid C45G7.";

RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA Waterston R.;

RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF067611; AAC19183.1;

DR InterPro: IPR002126; Cadherin.

DR PROSITE: PS50268; CADHERIN_2;

SQ SEQUENCE 986 AA; 109455 MW; 80729CDE64D9B93 CRC64;

Query Match 2.8%; Score 123.5; DB 5; Length 986;

Best Local Similarity 18.3%; Pred. No. 0.13;

Matches 128; Conservative 102; Mismatches 242; Indels 229; Gaps 32;

OY 127 WDVKAKSIGLELOFSIPR-----LRQIGFGESCPDG 158
 DB 387 FETKNNMAVACELHWSVPRKRSRQLECRLEKILQALRENNKRLLYLRGMHGKT-PVK 445
 OY 159 VTHSISGRIDATVVRIGTFCPSNCTVSRIRKQGVKALHLPWHPRVVVSFSTANSSIK 218
 DB 446 LADS-----STVENISKDRETSVKGITYORSYTR--LPYLAKETPEWHS----- 488
 OY 219 RLCTISVFEGESATLMSANY-----PEGFPD-----ELMTQFVVAHLRASVSFLNF 269
 DB 489 QYTTLOD---EKSDVESNRYPIQRPQKFLDTEPAELLITAFNSGNLTVSSITLTK 545
 OY 324 GI-LRQFQVLVQHQP-NBENKITYVVDLSNERAMSLTIERPVKOSKRVPGCFVLESR 381
 DB 606 NFNIDAEFEIYEPKVDPSKMTAEV-----CFIATSK 641
 OY 382 TC-SSNLTLSGSKH---KISFLCDDLRLMNVKRTISCTDHRQKRSYSLOVPSDI 436
 DB 642 NLDTSFVFPDANSKNKRVKVEFGD-----SVDSVANSGRINNGFSGNQSISSVGDV 694

Query Match	2.8%;	Score 123.5;	DB 6;	Length 3620;
Best Local Similarity	18.1%;	Pred. NO. 0.94;		
Matches 154;	Conservative 104;	Mismatches 285;	Indels 309;	Gaps 40

RESULT 14			
093VB7	PRELIMINARY:	PR7	724 AA.
AC	093VB7:		
DT	01-DEC-2001 (TREMBLrel. 19, Created)		
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)		
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)		
DE	P0487H02.1 protein (OSUBA0025P13.19 protein).		
GN	P0487H02.1.		
OS	Oryza sativa (Rice).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta		
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;		

OC Ehnhartioideae: Oryzeae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K. ;
 RT "Oryza sativa nippobare(Ga3) genomic DNA, chromosome 1, PAC
 clone: P0487H02. ";
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K. ;
 RT "Oryza sativa nippobare(Ga3) genomic DNA, chromosome 1, BAC
 clone: OSUBae0025P13. ";
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AP002883; BAB67842.1; -
 DR EMBL: AP003140; BAB5759.1; -
 DR InterPro: IPR001525; C5_DNA_meth.
 DR PROSITE: PS00094; C5_MTASE_1; UNKNOWN_1.
 DR PROSITE: PS00094; C5_MTASE_1; UNKNOWN_1.
 SQ SEQUENCE 724 AA; 79780 MW; DEB3E09BEAF23253 CRC64;

```

OY 348 DLNERNASLTIEPPY--KOSRKKVPCGFCVLEBRTSSNLTILSGSNKHSFLDDLT 405
Db 260 ELQREDACELGIGHNPILAADEREPP-----RSESELTTRSRISDILPSMADQLR 310
OY 406 RLW-----MNEKTI SC -TDHRYCQRKSYSLQVPSDILHLYELHDFSXKLLVPKD- 455
Db 311 RRMMAADGVLEYEVARRRIPCCEIDRRHMKRVCMSRMVAVLKLAAPRPPPLPYLALPESD 370
OY 456 ----RLSLVLPQKLOQHTHEKPCNT--SFSLVASALPISDLYGSGFCPG-----GSI 504
Db 371 GLPATSVCYL-----SGCRTAHSFSLQGA--RCARIGYSTIDGGGLFLAVGQ 415
OY 505 KOIQVKONISVT-----LRTAESPQOASRQGLVTSEIPIFYKE-----GVF 547
Db 416 AOCALLMLKINGEFTDLPLNLAIPVNSVNPNDREMAIYAATLSGCP--TEOGCIWAGII 473
OY 548 TVTTPD-----TKSKVLRTPNMORGSLSLTSVMSNIVSPRDVACILFFKERSGVQC 600
Db 474 ESSPLVAVGHVTRSIAR-----WRMGQVYVLPYLMAL--EEDNPLRLREVEDLICH 524
OY 601 TGRAFMIIQEORTRAEELISLDEVDLPKPRSFHNHSFWNINSCSPSGKODLFLSVTLT 660
Db 525 HG-AFHFL---TRAEDVLACE-----PRVFYRDSVSLVPRAN-----MFL 561
OY 661 PRVYD--LTVLLIAVAGGVLLLSALGLITICVKKKKKKTKNGRAVGIINGNINTEMPO 718
Db 562 PRVHDENETVLAARYLVGSKRL-----MVRLASGRGRTYS--AFRVF-----Q 605
OY 719 PKKFOGKRKDDSHVYAYIEDTMVYGHLLDSSGSLFLOPEVDYTRPPGTMVCP----- 773
Db 606 KKKFNTGEDEBPQSQRSAHFE--YWSLDELIDGMLLVGLGCSSTYKAGDORRYGMEBG 663
OY 774 -----PSPPLIC-----SAAPAKALTEE-----PPRSP 799
Db 664 VYFLDDPSIHOMITIGDAPKPYLCSNDKWSMAAPRPOGOVERCPRPERGPSIHSBP 719

```

RESULT	15
09Y216	
ID	09Y216
AC	09Y216; Q9NOE3;
DT	01-NOV-1999 (TrEMBLrel. 12, Created)
DT	01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE	KIAA0980 protein (Df59JIN24.1.1) (KIAA0980 protein, Isoform 1)
DE	(Fragment).
DE	KIAA0980 OR Df69JIN24.1.

OS	Homo sapiens (Human).
OC	Eumetazoa; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX	NCBT_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE-BRAIN;
RX	MEDLINE:999246063; PubMed=10231032;
RA	Nagase T., Ishikawa K., Suyama M., Kikuno R., Hirosewa M., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT	"Prediction of the coding sequences of unidentified human genes. XIII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro.";
RL	DNA Res. 6:63-70(1999).
RN	[2]
RP	SEQUENCE OF 22-1406 FROM N.A.
RA	Blakey S.;
RL	Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AB023197; BAA76824.1; -
DR	EMBL; AL031672; CAB99291.1; -
DR	HSSP; P02633; 4ICB.
DR	InferPro; IPR002048; EF-hand.
DR	Pfam; PF00036; ethand; 3.
DR	SMART; SM00054; EFh; 2.
DR	PROSITE; PS00018; EF_HAND; UNKNOWN_1.
FT	NON_TER
SQ	SEQUENCE 1406 AA; 158951 MW; A530020275943655 CRC64;

Query Match 2.7%; Score 118.5; DB 4; Length 1406;
 Best Local Similarity 17.8%; Fred. No. 0.61;
 Matches 154; Conservative 120; Mismatches 306; Indels 283; Gaps 37;

[illegible]


```
Db 763 AELSGE---LSGLGALPARDLTLELEPPQGPLPRGSGRSEQLLELRLKLOPCASEK 818
QY 694 KKKTKNGPAVGIIYNGNINTEMPROPKFKQGRKDN---DSHVYAVIEDTWYGHLLQD 749
Db 819 RQQMC-----VSIALEEFEELARGRKRVDPGLEEMQALPKDGLVAG----- 861
QY 750 SSGSEFLQPEVDYIRPQGTWVCPPSPPTICSRAPTAKL----- 788
Db 862 -SGQ-----EGTRGLPLLRPG--CGERPLAWLAPGDGSESEEAAGAPRRRQA 906
QY 789 ---ATEEPPRRSPPESE--SEPYTFSHP--NNGDVSSKD-----TDIP 824
Db 907 QDTEATQSPAPAPAPASHGSPSRWSRMOQCGVDGDIIVKPEPPFGASAAGLEQPGARELP 966
QY 825 LINTQE-----PMEPA 835
Db 967 LGTERDASQTOPRMWEPPLRPA 989
```

Search completed: November 11, 2002, 11:35:52
Job time : 88 secs

JOURNAL Patent: WO 0204508-A 3 17-JAN-2002;
FEATURES Location/Qualifiers
source 1..6163
/organism="Homo sapiens"
/db_xref="taxon:9606"
5'UTR 1..283
GC_signal 147..157
misc_feature 201..209
/note="cap signal; Transkriptionsstart"
283..2793
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD22954.1"
/db_xref="GI:18618679"
translation="MAGLNCGVSIALLGLVLLGARLPNGAEFAELAPRESNTVLI
KLGPTPLAKPCYIVISKRIITMSTIKSGRIYFECQSPENHEVLEKNIDMKG
PCPEVLOLPSTSLPTLNTFRFIMDKAKISIGLEOFSEIPRLROIQDPESPQVYT
HISIGRDAIVVATGTCFCSNGTYSRIKMQGVKMAHLHPFHRNNSGESIANSSIK
RLCTIESVEGESATLMSANYPEGEPEDELMTQEVPAHLRASVSFLPNLSNCR
KEERVEYIIGSTTNPVEFKLEDKQPGNMAENNLISQGDQDQSPGILRLQFVLY
OHPONESNKIYVDLSNERAMSLTIEPRPVKQSKSYLOVPSDILHLPELHPSMKL
SKKISFLCDLRLMMNVKTSCTDHYCQKRSYLAQPSDILHLPELHPSMKL
LVPRDLSTLVPAOKLOOHTHEKPCMTSFSYLYASAIPSDILYFSGSECGSISKOI
YKONISYTLTEFAPRQOEAROGRLTYSFTPYKEBEVFTYTDPKSKYLYLRPNMR
GLPSLVSNNISVPRDVACTFEKERSVVCOTGAENITIEORTAREITSLED
VLPKPSHHHSFKNWINSNCSPTSGKQDLLESVTLTPRTDVLVILAAVGGVLLS
ALGLIICCVKRRKTKNKPAVGIVNINITEMRPQKROKDKNDSHVAVIEDT
MYEGLHLDSSGSFLQPEVDYRPFQGTMGVCCPSPTICSRAPYAKLATEEPPRSP
PESESEYTFSSHNNNGDVSKDIDPLINQEPMEPAE"
3'UTR 2794..6163
BASE COUNT 1501 a 1689 c 1482 g 1491 t
ORIGIN

Query Match 100.0%; Score 6163; DB 6; Length 6163;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 6163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCAACGCCGCAATGGAGTAGTAGTGGAGACCCAGCAACCCGGTGCCGGAGCCCTGCACCC 60
DB 1 CCAACGCCGCAATGGAGTAGTAGTGGAGACCCAGCAACCCGGTGCCGGAGCCCTGCACCC 60
QY 61 TGGAGAGGAGAGGGGTCGCGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 61 TGGAGAGGAGAGGGGTCGCGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
QY 121 CGGGTCAGCTCGCATCTGCTGCGCAGAGGGGCTGCGGCGGCTCGCCGCGCGCGAG 180
DB 121 CGGGTCAGCTCGCATCTGCTGCGCAGAGGGGCTGCGGCGGCTCGCCGCGCGCGAG 180
QY 181 GTGAGTAGAGCCAGAGGCGGAGCGAGCTGCGCGGCTTGGGCGCTGGGGCCCGCGCTCC 240
DB 181 GTGAGTAGAGCCAGAGGCGGAGCGAGCTGCGCGGCTTGGGCGCTGGGGCCCGCGCTCC 240
QY 241 CCACCGCTGTTTCCCAACGAGGCGGAGGCGGAGGAGTCAGGCGGAGCTGAACGCG 300
DB 241 CCACCGCTGTTTCCCAACGAGGCGGAGGCGGAGGAGTCAGGCGGAGCTGAACGCG 300
QY 301 GGGGCTCTATCGCACTGCTAGGGGTTCTGCTGCTGGGTGCGGCGGCTGCGCGGCGG 360
DB 301 GGGGCTCTATCGCACTGCTAGGGGTTCTGCTGCTGGGTGCGGCGGCTGCGCGGCGG 360
QY 361 GCACAGAGCTTTGAGATTGCTCTCCACGAGAAAGCAACATTACAGTCTCTATAAGCTG 420
DB 361 GCACAGAGCTTTGAGATTGCTCTCCACGAGAAAGCAACATTACAGTCTCTATAAGCTG 420
QY 421 GGGAGCCCGAGACTGCTGGGAGAAACCGTGACATCGCATTTCTAAAGACATATAACC 480
DB 421 GGGAGCCCGAGACTGCTGGGAGAAACCGTGACATCGCATTTCTAAAGACATATAACC 480
QY 481 ATGTTGTTCATCAAGTCTGGAGAAAGATAGTCTTTACCTTAGCTGCACAGTCTGAG 540
DB 481 ATGTTGTTCATCAAGTCTGGAGAAAGATAGTCTTTACCTTAGCTGCACAGTCTGAG 540

QY 541 AATCACTTTGTCATAGAGATCCAGAAAAATATGACTGTATGTCAGGCCCATGTCCTTT 600
DB 541 AATCACTTTGTCATAGAGATCCAGAAAAATATGACTGTATGTCAGGCCCATGTCCTTT 600
QY 601 GGGAGAGTTGAGCTTCCAGCCCTCGACATCGTGTGGCTACCCCTCAAGAACTTCAATC 660
DB 601 GGGAGAGTTGAGCTTCCAGCCCTCGACATCGTGTGGCTACCCCTCAAGAACTTCAATC 660
QY 661 TGGAGTGCAGAGCTCATAGAGATCGGTTAGAGCTGACAGTTTCCATCCCTCGCCTG 720
DB 661 TGGAGTGCAGAGCTCATAGAGATCGGTTAGAGCTGACAGTTTCCATCCCTCGCCTG 720
QY 721 AGCAGATCGTCCGGGTGAGAGCTGCCAGAGGAGTCACTACTCATCAGCGGCCGA 780
DB 721 AGCAGATCGTCCGGGTGAGAGCTGCCAGAGGAGTCACTACTCATCAGCGGCCGA 780
QY 781 ATCATGCGACCGGTGTCAGATGAGAGACCTTGCACACATGGACATGTCGCCGGATC 840
DB 781 ATCATGCGACCGGTGTCAGATGAGAGACCTTGCACACATGGACATGTCGCCGGATC 840
QY 841 AAGATGCAGAGAGAGTGAATAATGGCTTACACCTCCCATGGTTCACACCCAGAAATGTC 900
DB 841 AAGATGCAGAGAGAGTGAATAATGGCTTACACCTCCCATGGTTCACACCCAGAAATGTC 900
QY 901 TCCGCTTCAGCATTCGCAACCCGCTCATCTATATAACGTCGTGTCATCATGAGTCTGTG 960
DB 901 TCCGCTTCAGCATTCGCAACCCGCTCATCTATATAACGTCGTGTCATCATGAGTCTGTG 960
QY 961 TTTGAGGGTGAAGGCTGCAACCCCTGATGTCGCAACCTCCAGAGAGGCTTCCCTGAG 1020
DB 961 TTTGAGGGTGAAGGCTGCAACCCCTGATGTCGCAACCTCCAGAGAGGCTTCCCTGAG 1020
QY 1021 GATGAGCTCATAGAGCTGAGCTTTGCTGCTCCTGCACACCTGCGGGCCAGGCTCTCTTC 1080
DB 1021 GATGAGCTCATAGAGCTGAGCTTTGCTGCTCCTGCACACCTGCGGGCCAGGCTCTCTTC 1080
QY 1081 CTCAACTTCAACCTCTCTCAACTGTGAGAGAGAGAGAGCGGTTGAATACTACATCCC 1140
DB 1081 CTCAACTTCAACCTCTCTCAACTGTGAGAGAGAGAGAGCGGTTGAATACTACATCCC 1140
QY 1141 GGCTTCACACCAACCCCGAGGAGTTCAGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
DB 1141 GGCTTCACACCAACCCCGAGGAGTTCAGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
QY 1201 GGGAACTTCAACCTCTCTCTCAAGGCTGTGACCAAGATGCCCAGAAATGTCAGGATCTCTC 1260
DB 1201 GGGAACTTCAACCTCTCTCTCAAGGCTGTGACCAAGATGCCCAGAAATGTCAGGATCTCTC 1260
QY 1261 CGGCTGCAAGTTCAGAGTTTGGTCCACATCCACAAAATGAAGCAATAAAATCTACGTG 1320
DB 1261 CGGCTGCAAGTTCAGAGTTTGGTCCACATCCACAAAATGAAGCAATAAAATCTACGTG 1320
QY 1321 GTTCACTTGAATGAGAGCGAGCATCTCAACATCGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
DB 1321 GTTCACTTGAATGAGAGCGAGCATCTCAACATCGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
QY 1381 AGCGCAAGTTTGTCCCTGCGCTGTGCTGTGTCTAGAAATCTCGAAGCTGAGTAGCAAC 1440
DB 1381 AGCGCAAGTTTGTCCCTGCGCTGTGCTGTGTCTAGAAATCTCGAAGCTGAGTAGCAAC 1440
QY 1441 CTCACCTGACATCTGGCTCCAAACACAAAATCTCTCTTTGTGATGATCTGACAGCT 1500
DB 1441 CTCACCTGACATCTGGCTCCAAACACAAAATCTCTCTTTGTGATGATCTGACAGCT 1500
QY 1501 CTGTGATGATGATGAGAAAAACCATAGCTGACAGAGACACAGGATCTGCAAGAGAA 1560
DB 1501 CTGTGATGATGATGAGAAAAACCATAGCTGACAGAGACACAGGATCTGCAAGAGAA 1560
QY 1561 TCTTACTACTCCAGGTTGCCAGAGTGAATCTCCACCTGCTGTGAGAGCTCATGACTTC 1620
DB 1561 TCTTACTACTCCAGGTTGCCAGAGTGAATCTCCACCTGCTGTGAGAGCTCATGACTTC 1620
QY 1621 TCTTGAAGGCTGTGTGTGCCAAGAGACAGGCTCAGCTGTGTGTGTGCCAGCCAGAG 1680

```
Db 1621 TCTGGAGGTGCTGGGTGGCCCAAGGAGACAGCTAGCCTGGTGGTGGCCAGGCCAAG 1680
Qy 1681 CTCACAGAGATACACACGAGAAAGCCCTGCACACACAGCTTCACATACCTGCGCCAGT 1740
Db 1681 CTCACAGAGATACACACGAGAAAGCCCTGCACACACAGCTTCACATACCTGCGCCAGT 1740
Qy 1741 GCCATGCCAGGAGGAGCAGTACTTGGCTCCTGCGCCGGGAGGCTCTATCAACAG 1800
Db 1741 GCCATGCCAGGAGGAGCAGTACTTGGCTCCTGCGCCGGGAGGCTCTATCAACAG 1800
Qy 1801 ATCCAGGTGAGAGAGAAACATCTGCTGACCCCTTCGACACTTTGGCCCACTTCACAA 1860
Db 1801 ATCCAGGTGAGAGAGAAACATCTGCTGACCCCTTCGACACTTTGGCCCACTTCACAA 1860
Qy 1861 GAGGCTCCAGGAGGCTGACGCTGCTTTATACCTTATTTCAAGAGAGGCGTT 1920
Db 1861 GAGGCTCCAGGAGGCTGACGCTGCTTTATACCTTATTTCAAGAGAGGCGTT 1920
Qy 1921 TTCAGGCTGACCCCTGACACAAAAGAGAGTCTACCTGAGAGCCCACTGGGAGCGG 1980
Db 1921 TTCAGGCTGACCCCTGACACAAAAGAGAGTCTACCTGAGAGCCCACTGGGAGCGG 1980
Qy 1981 GGCCTGCAATCCCTCAACCTCTGCTCTGAGACATCAGCGTGGCCAGAGACAGTGGCC 2040
Db 1981 GGCCTGCAATCCCTCAACCTCTGCTCTGAGACATCAGCGTGGCCAGAGACAGTGGCC 2040
Qy 2041 TGCCTGACTTTCTTTAAGAGCGGAGCGGCTGCTGCTCCAGACAGGCGCGCATTCATG 2100
Db 2041 TGCCTGACTTTCTTTAAGAGCGGAGCGGCTGCTGCTCCAGACAGGCGCGCATTCATG 2100
Qy 2101 ATCATCAGAGAGAGGAGCGGAGCGGCTGAGAGATCTTCAGCCCTGAGAGAGATGCTC 2160
Db 2101 ATCATCAGAGAGAGGAGCGGAGCGGCTGAGAGATCTTCAGCCCTGAGAGAGATGCTC 2160
Qy 2161 CCCAAGGCAAGCTTCCACATCAGAGCTTGGGTCAACATCTCTAAGCTCAGCCCAAG 2220
Db 2161 CCCAAGGCAAGCTTCCACATCAGAGCTTGGGTCAACATCTCTAAGCTCAGCCCAAG 2220
Qy 2221 AGGCGCAGAGAGCTAGACCTGCTCTCTGCGTGACACTTACCCAGAGAGCTGAGACTG 2280
Db 2221 AGGCGCAGAGAGCTAGACCTGCTCTCTGCGTGACACTTACCCAGAGAGCTGAGACTG 2280
Qy 2281 ACTGTCATCTCTATCGAGAGGCTGAGAGTGTACTGCTGCTGCTGCTGCTGCTGCTG 2340
Db 2281 ACTGTCATCTCTATCGAGAGGCTGAGAGTGTACTGCTGCTGCTGCTGCTGCTGCTG 2340
Qy 2341 ATCATTTGCTGTGTGAAAAGAGAAAAGAGACAAACAAAGGCGCCGCTGTGGTATC 2400
Db 2341 ATCATTTGCTGTGTGAAAAGAGAAAAGAGACAAACAAAGGCGCCGCTGTGGTATC 2400
Qy 2401 TACAAATGGCAACATCAATCTAGAGATGCGAGGAGCCCAAAAGTTTCAAGAAAGGCGA 2460
Db 2401 TACAAATGGCAACATCAATCTAGAGATGCGAGGAGCCCAAAAGTTTCAAGAAAGGCGA 2460
Qy 2461 AAGGACAAATGACCTCCATGCTATGAGCATGAGAGACATGATATGGGCACTG 2520
Db 2461 AAGGACAAATGACCTCCATGCTATGAGCATGAGAGACATGATATGGGCACTG 2520
Qy 2521 CTACAGGATTCACAGCGGCTCTTCCTGACAGCAGAGGTGAGACCTACCGCGCTTCAG 2580
Db 2521 CTACAGGATTCACAGCGGCTCTTCCTGACAGCAGAGGTGAGACCTACCGCGCTTCAG 2580
Qy 2581 GGCACCAATGGGGGTCTCTCTCTCCCTCCCAACCAACATATGCTGACAGGCGCCAACTGCA 2640
Db 2581 GGCACCAATGGGGGTCTCTCTCTCCCTCCCAACCAACATATGCTGACAGGCGCCAACTGCA 2640
Qy 2641 AAGTGGGCACTAGAGAGGAGCCCTGCTGCTCCCTCCCAACCAATATGCTGACAGGCGCCAACTGCA 2700
Db 2641 AAGTGGGCACTAGAGAGGAGCCCTGCTGCTCCCTCCCAACCAATATGCTGACAGGCGCCAACTGCA 2700
Qy 2701 ACCTTCTCCATCCCAACAAATGGGATGTAAAGAGCAAGAGACAGACATTCCTCTACTG 2760
Db 2701 ACCTTCTCCATCCCAACAAATGGGATGTAAAGAGCAAGAGACAGACATTCCTCTACTG 2760
Qy 2761 AACACTAGAGAGCCCATGAGAGCCAGAGAAATACCTGATCATTCACAGAGCTTGTCTGA 2820
Db 2761 AACACTAGAGAGCCCATGAGAGCCAGAGAAATACCTGATCATTCACAGAGCTTGTCTGA 2820
Qy 2821 GTTTCATAAAGCAGGAGCAGTACAGACCCCGCTGTTCCATCAACAGAAATCCTAAAGAA 2880
Db 2821 GTTTCATAAAGCAGGAGCAGTACAGACCCCGCTGTTCCATCAACAGAAATCCTAAAGAA 2880
Qy 2881 GAGGAATTAATACAGAGAGAGAGAGGTTTCTGAGACAGCCCAACTTCACATTCG 2940
Db 2881 GAGGAATTAATACAGAGAGAGAGAGGTTTCTGAGACAGCCCAACTTCACATTCG 2940
Qy 2941 TCACTGACATCTCTAAGGAGCAACATTTGAAATGATGATTCATCTGAGTACAGT 3000
Db 2941 TCACTGACATCTCTAAGGAGCAACATTTGAAATGATGATTCATCTGAGTACAGT 3000
Qy 3001 CATGACAGCTCATGCTGCTCCCAACTTAGGCTGTGCGGTTTGGCCAGCCTGTAAAGAGG 3060
Db 3001 CATGACAGCTCATGCTGCTCCCAACTTAGGCTGTGCGGTTTGGCCAGCCTGTAAAGAGG 3060
Qy 3061 AGAGAGCCTGAGTCACTAGACATAGGATGACAGAAAGCCCTGATTCAGAGTGTAAAC 3120
Db 3061 AGAGAGCCTGAGTCACTAGACATAGGATGACAGAAAGCCCTGATTCAGAGTGTAAAC 3120
Qy 3121 AGAGGCTTGCCTTTAGAGCAACAGTTCATTCAGAGAGCCTTACCTGAGTCCCTA 3180
Db 3121 AGAGGCTTGCCTTTAGAGCAACAGTTCATTCAGAGAGCCTTACCTGAGTCCCTA 3180
Qy 3181 CTCTCAGCTGGGCTCCAGAGTGAAGAGACAAATGTGCTTTTATTTATTTATTTG 3240
Db 3181 CTCTCAGCTGGGCTCCAGAGTGAAGAGACAAATGTGCTTTTATTTATTTATTTG 3240
Qy 3241 TGGTCTGTGTATTAAGAGATCAAAATGTATAACCACTAGCTCTTTCTACCTGACTTA 3300
Db 3241 TGGTCTGTGTATTAAGAGATCAAAATGTATAACCACTAGCTCTTTCTACCTGACTTA 3300
Qy 3301 GTATTAACCTATTAACCTAGTGTGATGCTGGGTGTGACTTCTACTGACCCCTGAT 3360
Db 3301 GTATTAACCTATTAACCTAGTGTGATGCTGGGTGTGACTTCTACTGACCCCTGAT 3360
Qy 3361 AAACGTGTGCTGCTCCCAAGAGTGTGGAATTAATTAACAATCTGTCCAACAGAAAGA 3420
Db 3361 AAACGTGTGCTGCTCCCAAGAGTGTGGAATTAATTAACAATCTGTCCAACAGAAAGA 3420
Qy 3421 ATGTGTGTGTTGAGACAGATTAACATATCTCTTGTATTAAGAGACTTCTGATTC 3480
Db 3421 ATGTGTGTGTTGAGACAGATTAACATATCTCTTGTATTAAGAGACTTCTGATTC 3480
Qy 3481 TAGTGTGCTGTGTTATCCATTTGGAATTAATTCATTCATTCCTATTAAGT 3540
Db 3481 TAGTGTGCTGTGTTATCCATTTGGAATTAATTCATTCATTCCTATTAAGT 3540
Qy 3541 CCTAGCAATAAGAGAAATTTCTCAAGTTTCCATGTGCGGTTCTCTAGCTGACAGATA 3600
Db 3541 CCTAGCAATAAGAGAAATTTCTCAAGTTTCCATGTGCGGTTCTCTAGCTGACAGATA 3600
Qy 3601 CTTTGCATTTAAAGAGAAATTTAGAGAAATTTCTCATCTCTTAAATGATA 3660
Db 3601 CTTTGCATTTAAAGAGAAATTTAGAGAAATTTCTCATCTCTTAAATGATA 3660
Qy 3661 TACCAAAACAGTGGCCCTGATTAAGTTTCTGTGTCACAGCAACCAATTAATGATG 3720
Db 3661 TACCAAAACAGTGGCCCTGATTAAGTTTCTGTGTCACAGCAACCAATTAATGATG 3720
Qy 3721 CTTTAAACAAACATTAAGTATAGTCTGGGATCAAGATTCCTCAAAATGATGATG 3780
Db 3721 CTTTAAACAAACATTAAGTATAGTCTGGGATCAAGATTCCTCAAAATGATGATG 3780
Qy 3781 GAATGAATAATCAAGGTGTACAGAGCTGTGCTCTCTTGAAGGCTCTAGGAGAAAGCCG 3840
Db 3781 GAATGAATAATCAAGGTGTACAGAGCTGTGCTCTCTTGAAGGCTCTAGGAGAAAGCCG 3840
```

QY 3841 GTTCCCTCCATTTCAGGAGGCTGCTGCTATTCCTCCAGGCTCCAGTGGCTG 3900
DB 3841 GTTCCCTCCATTTCAGGAGGCTGCTGCTATTCCTCCAGGCTCCAGTGGCTG 3900
QY 3901 AAGCTTTCTCAGATGAGATGATGATGATGATGATGATGATGATGATGATG 3960
DB 3901 AAGCTTTCTCAGATGAGATGATGATGATGATGATGATGATGATGATGATG 3960
QY 3961 AAAGCCACAGAGAGATGATGATGATGATGATGATGATGATGATGATGATG 4020
DB 3961 AAAGCCACAGAGAGATGATGATGATGATGATGATGATGATGATGATGATG 4020
QY 4021 AGAGCTTTTCCATGCAAGACATAGCAGAGTGGGATTAAGACAGAGATCTT 4080
DB 4021 AGAGCTTTTCCATGCAAGACATAGCAGAGTGGGATTAAGACAGAGATCTT 4080
QY 4081 TGGGCTGCTGATTTATTTCTCCATGACAGACCTTCCGCTGCTGCTGCTGCTGCT 4140
DB 4081 TGGGCTGCTGATTTATTTCTCCATGACAGACCTTCCGCTGCTGCTGCTGCTGCT 4140
QY 4141 CAAATGATCTGGGCAAGAGGATGTTTGTAGCTGGGACTCTAACTTAAAAA 4200
DB 4141 CAAATGATCTGGGCAAGAGGATGTTTGTAGCTGGGACTCTAACTTAAAAA 4200
QY 4201 ACCCCAGATCAGAAAGATCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4260
DB 4201 ACCCCAGATCAGAAAGATCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4260
QY 4261 TGGAGCTGGGACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4320
DB 4261 TGGAGCTGGGACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4320
QY 4321 CCACACTACGCTTATCTATCTCCCAATGAGGCTAAGTGTGTTGTTCTAGTATG 4380
DB 4321 CCACACTACGCTTATCTATCTCCCAATGAGGCTAAGTGTGTTGTTCTAGTATG 4380
QY 4381 CCCCTGCAAGTGTGCTTATTTGTAATGAAAAAGAAAGCTGGGATTAATCTTAATC 4440
DB 4381 CCCCTGCAAGTGTGCTTATTTGTAATGAAAAAGAAAGCTGGGATTAATCTTAATC 4440
QY 4441 TGAGTAGACATGAGACCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4500
DB 4441 TGAGTAGACATGAGACCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4500
QY 4501 TTTTCTTTTCTTTTAAATGAGACAGATCTCATCTGTTGCTGCTGCTGCTGCTGCT 4560
DB 4501 TTTTCTTTTCTTTTAAATGAGACAGATCTCATCTGTTGCTGCTGCTGCTGCTGCT 4560
QY 4561 TGGGCAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4620
DB 4561 TGGGCAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4620
QY 4621 AGCCCTCCAAATAGCTGGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4680
DB 4621 AGCCCTCCAAATAGCTGGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4680
QY 4681 TTTGAGAGACAGGATTTACCATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4740
DB 4681 TTTGAGAGACAGGATTTACCATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4740
QY 4741 AATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4800
DB 4741 AATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4800
QY 4801 CCCCAACCCATATTAATCAATTAATCAATTAATCAATTAATCAATTAATCAATTAAT 4860
DB 4801 CCCCAACCCATATTAATCAATTAATCAATTAATCAATTAATCAATTAATCAATTAAT 4860
QY 4861 CCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4920
DB 4861 CCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4920

QY 4921 CAATATTTCTGGGCAAAACTCTCTCAAAAAATTAATGTAATGTAATGTAATGTAATG 4980
DB 4921 CAATATTTCTGGGCAAAACTCTCTCAAAAAATTAATGTAATGTAATGTAATGTAATG 4980
QY 4981 ATTCACTGGGAAATGCAACCGCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5040
DB 4981 ATTCACTGGGAAATGCAACCGCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5040
QY 5041 TAGTCTCCAAATAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5100
DB 5041 TAGTCTCCAAATAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5100
QY 5101 TTAAGTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5160
DB 5101 TTAAGTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5160
QY 5161 CCAGTACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5220
DB 5161 CCAGTACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5220
QY 5221 CTGCGCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5280
DB 5221 CTGCGCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5280
QY 5281 CCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5340
DB 5281 CCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5340
QY 5341 ATTCCACCAAGTGGGATTTGGGCTTCTTAAAGCTGGGCTGCTGCTGCTGCTGCTGCT 5400
DB 5341 ATTCCACCAAGTGGGATTTGGGCTTCTTAAAGCTGGGCTGCTGCTGCTGCTGCTGCT 5400
QY 5401 CACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5460
DB 5401 CACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5460
QY 5461 TGGACAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5520
DB 5461 TGGACAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5520
QY 5521 CCAGCTGAGGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5580
DB 5521 CCAGCTGAGGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5580
QY 5581 TCAGGCTACGTAAGAGTGAAGTGCATTTGTTAAAGCCCTGCTGCTGCTGCTGCTGCT 5640
DB 5581 TCAGGCTACGTAAGAGTGAAGTGCATTTGTTAAAGCCCTGCTGCTGCTGCTGCTGCT 5640
QY 5641 GCTCATCCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5700
DB 5641 GCTCATCCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5700
QY 5701 AATGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5760
DB 5701 AATGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5760
QY 5761 CAAGCCACCGGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5820
DB 5761 CAAGCCACCGGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5820
QY 5821 GCTTCACTGGAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5880
DB 5821 GCTTCACTGGAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5880
QY 5881 CTGCTGCTTAAATTAACAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5940
DB 5881 CTGCTGCTTAAATTAACAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5940
QY 5941 CTACTTGTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6000
DB 5941 CTACTTGTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6000
QY 6001 AGCATGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6060

|||||
Db 6001 AGATGGCTCAACAGATCTGCTTAATTGGATGATACGAAAGCAGGCTTGGG 6060
Qy 6061 ATACAGATTCTTCTCTTCATTTGATGCCGTGACTGTGTGAAGAGATGTTTTGTCC 6120
|||||
Db 6061 ATACAGATTCTTCTCTTCATTTGATGCCGTGACTGTGTGAAGAGATGTTTTGTCC 6120
Qy 6121 GGAATATTAATTAATGCTGTGAGTCTGCCCAAAAAAAA 6163
|||||
Db 6121 GGAATATTAATTAATGCTGTGAGTCTGCCCAAAAAAAA 6163
RESULT 2
AY026461 5963 bp mRNA linear PRI 17-AUG-2001
LOCUS Homo sapiens CUB domain containing protein 1 (CDCP1) mRNA, complete
DEFINITION cds.
ACCESSION AY026461
VERSION AY026461.1 GI:14328878
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 5963)
Schertl-Mostagener, M., Sommergruber, W., Abseher, R., Hauptmann, R.,
Ambros, P. and Schmeißer, N.
IDENTIFICATION of a novel gene, CDCP1, overexpressed in human
colorectal cancer
Oncogene 20 (32), 4402-4408 (2001)
JOURNAL 11466621
MEDLINE 21359860
PUBMED 2 (bases 1 to 5963)
REFERENCE 2 (bases 1 to 5963)
Schertl-Mostagener, M., Sommergruber, W., Abseher, R., Hauptmann, R.,
Ambros, P. F. and Schmeißer, N.
TITLE Direct Submissions
JOURNAL Submitted (31-JAN-2001) Exploratory Research, Boehringer Ingelheim,
Austria, Dr. Boehringerasse 5-11, Vienna 1121, Austria
FEATURES
source location/Qualifiers
1..5963
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3p23-p21"
/cell_line="A549; Calu 6"
1..5963
/gene="CDCP1"
1..82
/gene="CDCP1"
83..2593
/gene="CDCP1"
/note="transmembrane protein"
/product="CUB domain containing protein 1"
/protein_id="AAK02058.1"
/db_xref="GI:14328878"
translation="MAGLNCGVSTALLGVLLGAARLRPGAAFEIALPRENTYVLI
KLGPTPLAKPCYIVISKRHITMISIKSERIVTFSCSPNNHVIETQKIDCMG
PCPGEVLDLPSTSLPTLPTLNRTFINDVKAHKSIGLSLDFISRLRIGEGSCPDGVT
HSISGRIDATVIRIGTSCNGTVSRKMEGVKMLHLHLPFRNVSIGSIARSISIK
RLCITIEVEEGSATLMSANPEGRPELMTWGFVPAHRAVSSEINFLMSNER
KEEREYIIPGSTTNPYFKLEDKPGNMGAFNLSLQCCDDASPGILRIQFOYLV
OHPNESNKITVVDLSNERAMSLTEPRVKOSKRFPECFVCLSRITSSNLTLSG
SKHKISFLCDDTLRLMNVETISCTDHYCKQKSYSLQVPSDILHLPELHDFSKL
LVKDRSLVLPAPKLODHHEKPCNTSFYLAVALPISODLYGSGFPGSISIKQI
VKONISVLRFPAPFOEASRQGLTSPFYEKEGVETVLPDRSKYVLRTPMMDR
GLPSLTFSMNISVNPDOYACLTFFKERSGVVCGRAFMIIQDRTAELEIFSLD
VLKPSFHHHSFVWISNCSPTSGKQDLLEVTLPRTVDLTVLIAVGGVLLS
ALGLITCCVKKKKKTKNGPAVGINNTMPQPKFKQGRKDNDSHVAIVEDT
NVYGLLDSSGSLQPEVDYRPFQGTMGVCPSPPTICSAAPAKLATEPPRSP
PESESEPYTFHSHPNNGDVSSKTDIPDLINTOPMEPAE"
2594..5963
/gene="CDCP1"

BASE COUNT 1463 a 1634 c 1392 g 1474 t
ORIGIN
Query Match 96.8%; Score 5963; DB 9; Length 5963;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5963; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 201 CGAGAGTGGCCGGGCTTGGGGCCGCTGGGCGCCGCTGCCACCGCGTTTCCACC 260
Db 1 CGAGAGTGGCCGGGCTTGGGGCCGCTGGGCGCCGCTGCCACCGCGTTTCCACC 60
Qy 261 GAGCGGAGGCGTCCGAGTCAATGGCGGCGCTGAAGTGGGGGTCTATTCGACACTGT 320
Db 61 GAGCGGAGGCGTCCGAGTCAATGGCGGCGCTGAAGTGGGGGTCTATTCGACACTGT 120
Qy 321 AGGGGTTCTGCTGCTGGGTGCGGCGGCGCTGCGCGCGGGGAGAAAGCTTTAGATTCG 380
Db 121 AGGGGTTCTGCTGCTGGGTGCGGCGGCGCTGCGCGCGGGGAGAAAGCTTTAGATTCG 180
Qy 381 TCTGCCACGAGAAAGCAATTCATGTTCTATTAAGCTGGGACCCGACTGCTGGC 440
Db 181 TCTGCCACGAGAAAGCAATTCATGTTCTATTAAGCTGGGACCCGACTGCTGGC 240
Qy 441 AAAACCTGTACATGCTCATTTCTAAAGACATTAACCATGTGTCATCAAGTCTGG 500
Db 241 AAAACCTGTACATGCTCATTTCTAAAGACATTAACCATGTGTCATCAAGTCTGG 300
Qy 501 AGAAGATATGCTTACCTTACCTTACCTGCGAGAGTCTGAGAAATCACTTTGTATAGAT 560
Db 301 AGAAGATATGCTTACCTTACCTTACCTGCGAGAGTCTGAGAAATCACTTTGTATAGAT 360
Qy 561 CCAAGAAATATATGACTGTATGTCAGGCCCATGTCTTTGGGAGGTTAGCTTCAGCC 620
Db 361 CCAAGAAATATATGACTGTATGTCAGGCCCATGTCTTTGGGAGGTTAGCTTCAGCC 420
Qy 621 CTCGACATCGTTGTGCTGCTGCTCAACAGAACTTTCATCTGGAGTCAACCTCATAA 680
Db 421 CTCGACATCGTTGTGCTGCTGCTCAACAGAACTTTCATCTGGAGTCAACCTCATAA 480
Qy 681 GAGCATCGTTTATAGCTGATGCTTCCATCCCTGCTGAGGAGATCGGTCGGGTGA 740
Db 481 GAGCATCGTTTATAGCTGATGCTTCCATCCCTGCTGAGGAGATCGGTCGGGTGA 540
Qy 741 GAGCTGCCGAGAGAGTCACTCATCCATCAGGGCGGAATGGATCCACCGGTGAG 800
Db 541 GAGCTGCCGAGAGAGTCACTCATCCATCAGGGCGGAATGGATCCACCGGTGAG 600
Qy 801 GATCGAACTTGTGACGAAATGCAATGCTGTCCCGGATCAAGTCAAGAAAGAGTGA 860
Db 601 GATCGAACTTGTGACGAAATGCAATGCTGTCCCGGATCAAGTCAAGAAAGAGTGA 660
Qy 861 AATGGCTTACACCTCCATGTTTCCACCCAGAAATGTCTCGGGCTTCAGATTCGAAA 920
Db 661 AATGGCTTACACCTCCATGTTTCCACCCAGAAATGTCTCGGGCTTCAGATTCGAAA 720
Qy 921 CCGCTCATCTATAAAGCTGTGTCATCATGAGTCTGTGTTGAGGGTAAAGGCTAGC 980
Db 721 CCGCTCATCTATAAAGCTGTGTCATCATGAGTCTGTGTTGAGGGTAAAGGCTAGC 780
Qy 981 AACCTGATCTGCGCAACTACCAAGAGGCTTCCGAGAGTACGCTCAAGCTGGCA 1040
Db 781 AACCTGATCTGCGCAACTACCAAGAGGCTTCCGAGAGTACGCTCAAGCTGGCA 840
Qy 1041 GTTGTGCTTCTGACACCTGGGGGAGCGTCTCTTCTCAACTCAACTCTTCCAA 1100
Db 841 GTTGTGCTTCTGACACCTGGGGGAGCGTCTCTTCTCAACTCAACTCTTCCAA 900
Qy 1101 CTGTGAGAGAGAGAGAGGCGGTGTAATCTACATCCGGGCTCCACCAACCCCGA 1160
Db 901 CTGTGAGAGAGAGAGAGGCGGTGTAATCTACATCCGGGCTCCACCAACCCCGA 960
Qy 1161 GGTGTCAAGCTGGAGAGCAAGCAGCTGGGAACATGGCGGGAACCTCAACTCTCTCT 1220
|||||

D 961 GGTGTTCAAGCTGAGGACAAAGCAAGCTGGGACATGGCGGGAACTTCAACTCTCTCT 1020
Q 1221 GCAAGGCTGTGACCAAGATGCCCAAAGTCCAGGGATCCTCCGGCTGCAGTTCCAAGTTT 1280
D 1021 GCAAGGCTGTGACCAAGATGCCCAAAGTCCAGGGATCCTCCGGCTGCAGTTCCAAGTTT 1080
Q 1281 GGTCCACATCCACAAAATGAAGCATTAATATCTACGTGGTTGACTTGAATGAGCG 1340
D 1081 GGTCCACATCCACAAAATGAAGCATTAATATCTACGTGGTTGACTTGAATGAGCG 1140
Q 1341 AGCCATGTCACATCAGTCGAGCCGCGTCAAAAGAGCGGCAAGTTGTCCTGG 1400
D 1141 AGCCATGTCACATCAGTCGAGCCGCGTCAAAAGAGCGGCAAGTTGTCCTGG 1200
Q 1401 CTGTTGCTGTGTAGAAATCTCGACCTGCAGTACCACTGCATCTGAGCTC 1460
D 1201 CTGTTGCTGTGTAGAAATCTCGACCTGCAGTACCACTGCATCTGAGCTC 1260
Q 1461 CAACACAAAAATCTCTCTCTCTTGTATGATCTGACACGTCGTGGATGAATGGA 1520
D 1261 CAACACAAAAATCTCTCTCTCTTGTATGATCTGACACGTCGTGGATGAATGGA 1320
Q 1521 AACCTAAAGCTGACAGACCAAGCTGCTCAAAAGGAAATCTACTACCTCCAGTGGC 1580
D 1321 AACCTAAAGCTGACAGACCAAGCTGCTCAAAAGGAAATCTACTACCTCCAGTGGC 1380
Q 1581 CAGTGACATCTCCAGCTGCTGTGAGAGCTGATGATCTGACACGTCGTGGATGA 1640
D 1381 CAGTGACATCTCCAGCTGCTGTGAGAGCTGATGATCTGACACGTCGTGGATGA 1440
Q 1641 CAAGGACAGGCTCAGCTGCTGTGAGAGCTGATGATCTGACACGTCGTGGATGA 1700
D 1441 CAAGGACAGGCTCAGCTGCTGTGAGAGCTGATGATCTGACACGTCGTGGATGA 1500
Q 1701 GAAAGCTGCAACACAGCTCAGCTGATGATCTGACACGTCGTGGATGA 1760
D 1501 GAAAGCTGCAACACAGCTCAGCTGATGATCTGACACGTCGTGGATGA 1560
Q 1761 GTACTTGGGCTCTCTGCGCGGAGGCTCTATCAAGCAGATCCAGGTGAAGCAACAT 1820
D 1561 GTACTTGGGCTCTCTGCGCGGAGGCTCTATCAAGCAGATCCAGGTGAAGCAACAT 1620
Q 1821 CTCGGTACCTTGGACCTTGGCCCAAGCTTCCAAAGAGGCTTCCAGGAGGCT 1880
D 1621 CTCGGTACCTTGGACCTTGGCCCAAGCTTCCAAAGAGGCTTCCAGGAGGCT 1680
Q 1881 GACGGTCTCTTATACCTTATTCAAAGAGGCTTTCACGGTGACCTTGACAC 1940
D 1681 GACGGTCTCTTATACCTTATTCAAAGAGGCTTTCACGGTGACCTTGACAC 1740
Q 1941 AAAAAAGAGGTCTACGTGAGAGCCCAACTGGGACCGGGGCTGCCATCCTCACCTC 2000
D 1741 AAAAAAGAGGTCTACGTGAGAGCCCAACTGGGACCGGGGCTGCCATCCTCACCTC 1800
Q 2001 TGTGTCTGTGAACATCAAGCGGCGGAGAGACAGAGTGGCTGCTCTTAAAGA 2060
D 1801 TGTGTCTGTGAACATCAAGCGGCGGAGAGACAGAGTGGCTGCTCTTAAAGA 1860
Q 2061 GCGGAGGCGGTGTGCTGCGAGACAGGCGGCTATCATGATCATCCAGAGAGCGGAC 2120
D 1861 GCGGAGGCGGTGTGCTGCGAGAGCGGCGGCTATCATGATCATCCAGAGAGCGGAC 1920
Q 2121 CCGGGCTGAGAGATCTTCAAGCTGAGAGAGATGTCTCCCAAGCCAAAGCTTCCACA 2180
D 1921 CCGGGCTGAGAGATCTTCAAGCTGAGAGAGATGTCTCCCAAGCCAAAGCTTCCACA 1980
Q 2181 TCACAGCTTCTGGGTCAACATCTTAAGTGAAGCCCAAGCGGCAAGAGCTAGACT 2240
D 1981 TCACAGCTTCTGGGTCAACATCTTAAGTGAAGCCCAAGCGGCAAGAGCTAGACT 2040
Q 2241 GCTCTTCTGCGGTGACCTTACCCCAAGAGCTGAGCTTGAATCCATCCGAGG 2300
D 2041 GCTCTTCTGCGGTGACCTTACCCCAAGAGCTGAGCTTGAATCCATCCGAGG 2100
Q 2301 GGTGGAGGTGAGTCTTACTGCTGTGCTGCCCTGAGCTCATCTTGTGTA 2360
D 2101 GGTGGAGGTGAGTCTTACTGCTGTGCTGCCCTGAGCTCATCTTGTGTA 2160
Q 2361 GAAAGAAAAAGAACAAAGAGGCGCGCTGTGGATCTCAATGCAATCAATAC 2420
D 2161 GAAAGAAAAAGAACAAAGAGGCGCGCTGTGGATCTCAATGCAATCAATAC 2220
Q 2421 TGAGATGCGGAGGAGCCAAAAAGTTTCAAGAGGCGGAAAGCAATGACTCCATGT 2480
D 2221 TGAGATGCGGAGGAGCCAAAAAGTTTCAAGAGGCGGAAAGCAATGACTCCATGT 2280
Q 2481 GTATGATCATCAGAGACACCATGTATATGAGCATCTGCTACAGATTCAGAGGCTC 2540
D 2281 GTATGATCATCAGAGACACCATGTATATGAGCATCTGCTACAGATTCAGAGGCTC 2340
Q 2541 CTTCCTGACGACAGAGTGGACACCTTACAGGCGCATGAGGCTGTCTC 2600
D 2341 CTTCCTGACGACAGAGTGGACACCTTACAGGCGCATGAGGCTGTCTC 2400
Q 2601 TCCCTCCCAACACCATATGCTCCAGGCGCCCAACTGCAAGTTGGCAGTGAAGAGCC 2660
D 2401 TCCCTCCCAACACCATATGCTCCAGGCGCCCAACTGCAAGTTGGCAGTGAAGAGCC 2460
Q 2661 ACCTCTGCTCCCTCTCTGAGTCTGAGAGTGAACCGTACACTTCTCCATCCACAA 2720
D 2461 ACCTCTGCTCCCTCTCTGAGTCTGAGAGTGAACCGTACACTTCTCCATCCACAA 2520
Q 2721 TGGGATGTAGAGCAAGAGACACAGACATCCCTTACTGAACTCGAGGCGCATGGA 2780
D 2521 TGGGATGTAGAGCAAGAGACACAGACATCCCTTACTGAACTCGAGGCGCATGGA 2580
Q 2781 GCCAGCAAAATTAATCTGATCTCAGAGCTTGTGATGATTTCAATTAAGAGGCACT 2840
D 2581 GCCAGCAAAATTAATCTGATCTCAGAGCTTGTGATGATTTCAATTAAGAGGCACT 2640
Q 2841 GAGACACCGCTGCTGTTCTTAACAGAAATCTTAAGAGAGAAATTAACAGAGAA 2900
D 2641 GAGACACCGCTGCTGTTCTTAACAGAAATCTTAAGAGAGAAATTAACAGAGAA 2700
Q 2901 CAGAGAGGTTTCTGAGACAGCGGCAACTTCAATGCTGATGAGCTATCTAAG 2960
D 2701 CAGAGAGGTTTCTGAGACAGCGGCAACTTCAATGCTGATGAGCTATCTAAG 2760
Q 2961 GCAAGACATTAATGATGATTAATCAATCGATGATGATGATGATGATGATGATGAT 3020
D 2761 GCAAGACATTAATGATGATTAATCAATCGATGATGATGATGATGATGATGATGAT 2820
Q 3021 TCAACTTAAGCTGTGCGGTTAGCCAGCTGTATGAGAGAGAGGCTGATCACCTA 3080
D 2821 TCAACTTAAGCTGTGCGGTTAGCCAGCTGTATGAGAGAGAGGCTGATCACCTA 2880
Q 3081 GCATAGGTTGCAAGAGCCCTGATTCAGAGTTTAAACAGAGGCTGCTTCCAG 3140
D 2881 GCATAGGTTGCAAGAGCCCTGATTCAGAGTTTAAACAGAGGCTGCTTCCAG 2940
Q 3141 ACAACAGTTCATTCAGAGGCTTACCTGAGTCCCTACTGAGGCTGCCAGG 3200
D 2941 ACAACAGTTCATTCAGAGGCTTACCTGAGTCCCTACTGAGGCTGCCAGG 3000
Q 3201 ATGAAAAGCAATGTGCTTTTATATATTTATTTGAGGCTGCTGTTATTAAGA 3260
D 3001 ATGAAAAGCAATGTGCTTTTATATATTTATTTGAGGCTGCTGTTATTAAGA 3060
Q 3261 GATTAATGATTAACCAACATCTTATCTTCACTGATTAATTAATTAATTAATTA 3320
D 3061 GATTAATGATTAACCAACATCTTATCTTCACTGATTAATTAATTAATTAATTA 3120
Q 3321 GTTGGATGCTGAGGTTGTGATCTTACTGACCGCTGATTAAGCTGTGCTGCCCA 3380
D 3121 GTTGGATGCTGAGGTTGTGATCTTACTGACCGCTGATTAAGCTGTGCTGCCCA 3180

QY	3381	GGTGGTGGGAATTAATTACAACTGTGCCAACCAGAAAGAAATGTGTGTGTTGAGCACA	3440
Db	3181	GGTGGTGGGAATTAATTACAACTGTGCCAACCAGAAAGAAATGTGTGTGTTGAGCACA	3240
QY	3441	TTGACACATATTCAGCTTTGATTAAGACACTTCCGTGATTCGTAGGTCGGTGTGGTATTC	3500
Db	3241	TTGACACATATTCAGCTTTGATTAAGACACTTCCGTGATTCGTAGGTCGGTGTGGTATTC	3300
QY	3501	CCATTTGTGAAATTCATCTTGAATCCCATTTGTCTATAGTCTTAGCAATTAAGAAATTT	3560
Db	3301	CCATTTGTGAAATTCATCTTGAATCCCATTTGTCTATAGTCTTAGCAATTAAGAAATTT	3360
QY	3681	CATTAGTTTTCTGTGGCACTGCACGCCATTTACTTGTGTAAATAAACAACACATTAGC	3740
Db	3481	CATTAGTTTTCTGTGGCACTGCACGCCATTTACTTGTGTAAATAAACAACACATTAGC	3540
QY	3741	TTATATGCTCGGGGATCAGAAATTCCAAATTCGATGTCCTGTAGTAAATCAAGGTGCA	3800
Db	3541	TTATATGCTCGGGGATCAGAAATTCCAAATTCGATGTCCTGTAGTAAATCAAGGTGCA	3600
QY	3801	GCACAGCTGTGTCCTCTTCTGTAAGAGGCTCTAGGGAGAAAGCGGTCTCCATTCACAGCT	3860
Db	3601	GCACAGCTGTGTCCTCTTCTGTAAGAGGCTCTAGGGAGAAAGCGGTCTCCATTCACAGCT	3660
QY	3861	TCTAGAGGCTGGCTGCATTCGCCAGGCTCCAGTGGCTGATCAGCTTTTCTCACATGGCAT	3920
Db	3661	TCTAGAGGCTGGCTGCATTCGCCAGGCTCCAGTGGCTGATCAGCTTTTCTCACATGGCAT	3720
QY	3921	CACGTGTACACTGGCCCTCCACTTCCCTCTTGTGACTTACAAAGCCACACAGAAATCC	3980
Db	3721	CACGTGTACACTGGCCCTCCACTTCCCTCTTGTGACTTACAAAGCCACACAGAAATCC	3780
QY	3981	AGGATATCTCTCATCTCTAAAGATCCTTATCATCTCTGTAAGAGCCTTTGCCATGCAAG	4040
Db	3781	AGGATATCTCTCATCTCTAAAGATCCTTATCATCTCTGTAAGAGCCTTTGCCATGCAAG	3840
QY	4041	ACAACATAGCACAGGTGGGATTTAGAGCACAGACATCTTGGGTGGCTGTATTATTCGCC	4100
Db	3841	ACAACATAGCACAGGTGGGATTTAGAGCACAGACATCTTGGGTGGCTGTATTATTCGCC	3900
QY	4101	TACCACACCTCTCTGCCACTGACTCCACAGAGAGGCTACAAAATGATCTGGCGCACAG	4160
Db	3901	TACCACACCTCTCTGCCACTGACTCCACAGAGAGGCTACAAAATGATCTGGCGCACAG	3960
QY	4161	GGATGTTTTTTAGCTTGGGACTTAACACTTAAAAAAACCCAGATCAAGAAATCTG	4220
Db	3961	GGATGTTTTTTAGCTTGGGACTTAACACTTAAAAAAACCCAGATCAAGAAATCTG	4020
QY	4221	GCCATGCTGGGGCTCAATCTCAGCTACCAACAACCTGGCTGGAGCTGGGACACAGGCT	4280
Db	4021	GCCATGCTGGGGCTCAATCTCAGCTACCAACAACCTGGCTGGAGCTGGGACACAGGCT	4080
QY	4281	GCCCTTAAAGAGGGGTCTCACTTACACAGGTACCAACAGCCACACTAGCCCTATCACT	4340
Db	4081	GCCCTTAAAGAGGGGTCTCACTTACACAGGTACCAACAGCCACACTAGCCCTATCACT	4140
QY	4341	TCCCAACAATAGGCTAAGTGTGTTTCTTACATGATCAATGCCCTGCAGAGTTGCAATTTAT	4400
Db	4141	TCCCAACAATAGGCTAAGTGTGTTTCTTACATGATCAATGCCCTGCAGAGTTGCAATTTAT	4200
QY	4401	TGTAAATGAAAAAAGAAAGACAGGGGATTAATCTGTATTCAGAGTGAAGACATGAGACCAA	4460
Db	4201	TGTAAATGAAAAAAGAAAGACAGGGGATTAATCTGTATTCAGAGTGAAGACATGAGACCAA	4260
QY	4461	TGTGTGTCTACATTAACCTTTTCTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTATG	4520

[illegible]

Db 5341 TCTTTTCTTTAAAGCCAGCCCACTTACTTAATACTTTCAGGCTCAGTGAAGAACT 5400
QY 5601 GAAGTGCATTTTGTGAAGCCTACTGCATGCAGCCACCTGCTCATCCAGCTGGTCTCC 5660
Db 5401 GAAGTGCATTTTGTGAAGCCTACTGCATGCAGCCACCTGCTCATCCAGCTGGTCTCC 5460
QY 5661 ATGCTTACGAGGAAGGCGAGGCGATGCAGAGCTGCTCTATGCTGTGTGATTTGACA 5720
Db 5461 ATGCTTACGAGGAAGGCGAGGCGATGCAGAGCTGCTCTATGCTGTGTGATTTGACA 5520
QY 5721 GAAGGGAAGGCTCAAGGAAGAGTCAACTGGGCAAGCAAGCCACCGAGCATGGCC 5780
Db 5521 GAAGGGAAGGCTCAAGGAAGAGTCAACTGGGCAAGCAAGCCACCGAGCATGGCC 5580
QY 5781 TTGGTAAAGGTAGCAGAGCTGGTGTGTGATCTGCAGTCTCAGTGAATAATTTA 5840
Db 5581 TTGGTAAAGGTAGCAGAGCTGGTGTGTGATCTGCAGTCTCAGTGAATAATTTA 5640
QY 5841 TTCAATTGAGATTTTATTTAGTGGCATTTTATTCATTTCTGTGCTTTAATAACAA 5900
Db 5641 TTCAATTGAGATTTTATTTAGTGGCATTTTATTCATTTCTGTGCTTTAATAACAA 5700
QY 5901 TGTACCAAAAAACAAGTATCAAGCTGTTTAAGTCTCGGCTACTGTGCTGGTTCAG 5960
Db 5701 TGTACCAAAAAACAAGTATCAAGCTGTTTAAGTCTCGGCTACTGTGCTGGTTCAG 5760
QY 5961 TAGAGGCCCGCGTTCCAGTGTGTGATCTGCAGAGGCTGCAGATGGGCTGCAGAGATG 6020
Db 5761 TAGAGGCCCGCGTTCCAGTGTGTGATCTGCAGAGGCTGCAGATGGGCTGCAGAGATG 5820
QY 6021 TGTCTTAAATTTGTGATGATACAGAAAGCCAGGCTTTGGATACAGTTCTTCTCTTC 6080
Db 5821 TGTCTTAAATTTGTGATGATACAGAAAGCCAGGCTTTGGATACAGTTCTTCTCTTC 5880
QY 6081 ATTGATGCGCTGCTGCTGTGTGAGAGAGATGTTTTCGCGAATAAATAATAGTCT 6140
Db 5881 ATTGATGCGCTGCTGCTGTGTGAGAGAGATGTTTTCGCGAATAAATAATAGTCT 5940
QY 6141 TGGAGTCTCGCCAAAAA 6163
Db 5941 TGGAGTCTCGCCAAAAA 5963

RESULT 3
AX353603 5897 bp DNA linear PAT 06-FEB-2002
LOCUS AX353603
DEFINITION Sequence 1 from Patent WO0204508.
ACCESSION AX353603
VERSION AX353603.1 GI:18618676
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Schwaefler, N., Scherl, M., Mostagaer, M., Sommergruber, W., and Abseher, R.
TITLE Tumour-associated antigen (b545), characterised by an amino acid
sequence as in seq. id. No. 4
JOURNAL Patent: WO 0204508-A 1 17-JAN-2002;
FEATURES
source location/Qualifiers
1. 5897
/organism="Homo sapiens"
/db_xref="taxon:9606"
1. 214
215. 2464
/note="unamed protein product"
/codon_start=1
/protein_id="CAD22953.1"
/db_xref="GI:18618676"
/translation="MLSIKSGRIIVTFSCSPENHVFIEIKNIDOMSGPCEGYO
LOPSTSLPTLNTFIMDYAKHSIGLEIOPSLPRLROIQPGSPGPDGYHRSISGRID
ATYVRIGTCSNCTVSRIKQEGYKALHLPHPHKPVSVFSSTANSSIKRLITISY
FEGESATILMSANYPGPFEDILTWOFVPAHLRAVSFSLNLSNCEKREVEREY

IPGSTNPEVFKLEDKOPGMAGNFNLSLOGCDODASPEILKOFVLYOHPOESN
KIIVYDLISNERAMSLTIERPVRKSRFVCEVCELSRLTJTSKRIISPL
CDLITRLMNVNERTISCTDHRKORXYSLOVSDILHLEVEHDFEMKLLVPRDLS
LVLEPAOKLOHHEKPCNTFSYLVAALPISODLYGSCPGSISIOIYKONISVT
LRTFAPSPFOEASROGLIVSFIRYFREGEVFTYPTDPSKSVYLRTPMDGLSLTSV
SNINISPRDVAULTPEKERSGVVCCGRAPMIIIOBRTABEFLSDEVLPRPSH
HSHWNIISGSPISGKODLPLSVLTPPTVDITVLLIIVAGGVLLTSLALGILIC
VKRRKKTNRKPAVGIYNGNINTEMPSQKSFREGEFTMTPMKQSSRTWYMGICR
IPAPSCSQWPTPTGRSRAPWGSVLPHPYAPGPOLOSPMLSLPLSLRVNT
PSPIDPMGM"
2465. 5897
BASE COUNT 1464 a 1596 c 1364 g 1467 t 6 others
ORIGIN
Query Match 95.0%; Score 5854.2; DB 6; Length 5897;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 5865; Conservative 2; Mismatches 5; Indels 1; Gaps 1;
QY 291 CCTGAACCTGGGGGTCTCTATCGACAGTCTGAGGCTTCTGCTGGTGGCGGCGCT 350
Db 25 CCTGAACCTGGGGGTCTCTATCGACAGTCTGAGGCTTCTGCTGGTGGCGGCGCT 84
QY 351 GCGCGGGGGGAGAGGCTTTGAGATTGCTCTGCGACAGAGAAAGCAATTACAGTCT 410
Db 85 GCGCGGGGGGAGAGGCTTTGAGATTGCTCTGCGACAGAGAAAGCAATTACAGTCT 144
QY 411 CATAAAGCTGGGGAGCCCGACTGCTGGCAAAACCTGTATCATCTGATTTCTAAAG 470
Db 145 CATAAAGCTGGGGAGCCCGACTGCTGGCAAAACCTGTATCATCTGATTTCTAAAG 204
QY 471 ACATATAACATGTTGTCATCAAGTCTGAGAAAGAAATAGTCTTTTACCTTTAGCTGCA 530
Db 205 ACATATAACATGTTGTCATCAAGTCTGAGAAAGAAATAGTCTTTTACCTTTAGCTGCA 264
QY 531 GAGTCCGAGATACATTTGTCATAGAGATCCAGAAATAATGACGTATGTCAGGCCC 590
Db 265 GAGTCCGAGATACATTTGTCATAGAGATCCAGAAATAATGACGTATGTCAGGCCC 324
QY 591 ATGTCTTTTGGGAGGTTGAGCTTCAAGCCCTGACATCGTTGTTGCTACCTCAACAG 650
Db 325 ATGTCTTTTGGGAGGTTGAGCTTCAAGCCCTGACATCGTTGTTGCTACCTCAACAG 384
QY 651 AACTTTCATCTGGGATGTCAAAGCTCATAGAGATGGGTTTAAAGCTGCAAGTTTCAT 710
Db 385 AACTTTCATCTGGGATGTCAAAGCTCATAGAGATGGGTTTAAAGCTGCAAGTTTCAT 444
QY 711 CCTCGGCTGAGGAGATCGGTCCGGGTGAGAGCTGCCACAGAGACTCACTCATCAT 770
Db 445 CCTCGGCTGAGGAGATCGGTCCGGGTGAGAGCTGCCACAGAGACTCACTCATCAT 504
QY 771 CAGCGGCGGAATCGATGCCACCGTGTGAGAGATCGGAACCTTTCGACGAATGCGACTGT 830
Db 505 CAGCGGCGGAATCGATGCCACCGTGTGAGAGATCGGAACCTTTCGACGAATGCGACTGT 564
QY 831 GTCCCGATCAAGATGCAAGAGAGGAAATAGGCTTACACTCCCATGCTTCCACCC 890
Db 565 GTCCCGATCAAGATGCAAGAGAGGAAATAGGCTTACACTCCCATGCTTCCACCC 624
QY 891 CAGAAATGTCTCCGGCTTACAGATTCGAAACCGCTCATCTTAAAGCTGTGCATCAT 950
Db 625 CAGAAATGTCTCCGGCTTACAGATTCGAAACCGCTCATCTTAAAGCTGTGCATCAT 684
QY 951 CGAGTCTGTGTTGAGGGTGAAGGCTGACGAACCCGTATGTCCTCAACTACCGAGAAG 1010
Db 685 CGAGTCTGTGTTGAGGGTGAAGGCTGACGAACCCGTATGTCCTCAACTACCGAGAAG 744
QY 1011 CTTCCTGAGAGATGATCATGACGTGCGAGTGTGCTTCTGACACCTGCGGCGCAG 1070
Db 745 CTTCCTGAGAGATGATCATGACGTGCGAGTGTGCTTCTGACACCTGCGGCGCAG 804
QY 1071 CGTCTCTTCTCTCACTTCAACCTCTCAACTGTGAGAGAGAGAGAGGCGGTTGAATA 1130
Db 805 CGTCTCTTCTCTCACTTCAACCTCTCAACTGTGAGAGAGAGAGAGGCGGTTGAATA 864

Qy	1131	CTAATCCGGGGCTCCACCAACCAACCCGAGGTGTTCAAAGTGGAGGACAAAGCAGCTGG	1130
Db	865	CTACATCCCGGGGTCCACACCACCAACCCCGAGGGTTTCAAAGCTGGAGGACAAAGCAGCTGG	924
Qy	1191	GAACATGGGGGAAACCTCAACCTCTCTGCAAGGCTGTGACCAAGATGCCAAAGTCC	1250
Db	925	GAACATGGGGGAAACCTCAACCTCTCTGCAAGGCTGTGACCAAGATGCCAAAGTCC	984
Qy	1251	AGGGATCTCCGGCTGACAGTTCCTCAAAATTTGGTCCAACTCCACAAATGAAAGCAATAA	1310
Db	985	AGGGATCTCCGGCTGACAGTTCCTCAAAATTTGGTCCAACTCCACAAATGAAAGCAATAA	1044
Qy	1311	AATCTACGTGGTGAATTGAGTGAAGGAGCCATGTCTACTACACCATCGAGCCAGCGCC	1370
Db	1045	AATCTACGTGGTGAATTGAGTGAAGGAGCCATGTCTACTACACCATCGAGCCAGCGCC	1104
Qy	1371	CGTCAACAGAGCCGCAAGTTTGTCCCTGGCTGTTTCGTGTCTTGAATCTCGAGCTG	1430
Db	1105	CGTCAACAGAGCCGCAAGTTTGTCCCTGGCTGTTTCGTGTCTTGAATCTCGAGCTG	1154
Qy	1431	CAGTAGCAACCTCAACCTGTACATCTGGGTCCAAACCAAAATCTCTTCCTTGGATGA	1490
Db	1165	CAGTAGCAACCTCAACCTGTACATCTGGGTCCAAACCAAAATCTCTTCCTTGGATGA	1224
Qy	1491	TCGTACACGTCTGTGATGAATGGAAAAAACAATAGCTGCACAGACCAACCGGTACTG	1550
Db	1225	TCGTACACGTCTGTGATGAATGGAAAAAACAATAGCTGCACAGACCAACCGGTACTG	1284
Qy	1551	CCAAAGAAATCTCTACTCACTCCAGGTGCCAGTGCATCTCTCCACCTGCTGTGAGCT	1610
Db	1285	CCAAAGAAATCTCTACTCACTCCAGGTGCCAGTGCATCTCTCCACCTGCTGTGAGCT	1344
Qy	1611	GCATGACTTCTCCGGAAGTGTGTGTGCCCAAAGACAGCTCACCCTGGTGTGTGCC	1670
Db	1345	GCATGACTTCTCCGGAAGTGTGTGTGCCCAAAGACAGCTCACCCTGGTGTGTGCC	1404
Qy	1671	AGCCCAACACCTGCAGCAGCATACACAGAAAGCCCTCAACACACACTTCAGTACTCT	1730
Db	1405	AGCCCAACACCTGCAGCAGCATACACAGAAAGCCCTGCACACACACTTCAGTACTCT	1484
Qy	1731	CGTGGCAGTGCCATACCCAGCCAGGACCTGTACTTGGGCTCTTCTGCCCCGAGGCTC	1790
Db	1465	CGTGGCAGTGCCATACCCAGCCAGGACCTGTACTTGGGCTCTTCTGCCCCGAGGCTC	1524
Qy	1791	TATCAACCAATCCAGTGAAGAGAAACATCTCGGTGACCTTCGACACTTGGCCCCAG	1850
Db	1525	TATCAACCAATCCAGTGAAGAGAAACATCTCGGTGACCTTCGACACTTGGCCCCAG	1584
Qy	1851	CTTCAACAAGAGAGCCCTCCAGGACGGGTGACAGGGTCTCTTATACCTTATTTCAAGA	1910
Db	1585	CTTCAACAAGAGAGCCCTCCAGGACGGGTGACAGGGTCTCTTATACCTTATTTCAAGA	1644
Qy	1911	GGAAGCGTTTTCACGGTGAACCCCTGACACAAAAAGCAAGGTCTACTGAGGACCCCCA	1970
Db	1645	GGAAGCGTTTTCACGGTGAACCCCTGACACAAAAAGCAAGGTCTACTGAGGACCCCCA	1704
Qy	1971	CTGGGACCGGGGCTGCTCACTCTCTGCTGTCTGTGAACATACAGCTGCCAGAGA	2030
Db	1705	CTGGGACCGGGGCTGCTCACTCTCTGCTGTGTGAACATACAGCTGCCAGAGA	1764
Qy	2031	CCAGGTGGCTGCTGACTTCTTTTAAAGAGGGGAGCGCGGTGCTGACAGACGGGCG	2090
Db	1765	CCAGGTGGCTGCTGACTTCTTTTAAAGAGGGGAGCGCGGTGCTGACAGACGGGCG	1824
Qy	2091	CGCATTCATGATCATCCAGAGACGCGGACCCGGGCTGAGAGATCTTCACACTGAGAGA	2150
Db	1825	CGCATTCATGATCATCCAGAGACGCGGACCCGGGCTGAGAGATCTTCACACTGAGAGA	1884
Qy	2151	GGATGTGCTCCCAAGCAAGCTTCCACATACAGCTTCTGGGTCAACATCTCTAACTG	2210
Db	1885	GGATGTGCTCCCAAGCAAGCTTCCACATACAGCTTCTGGGTCAACATCTCTAACTG	1944

OY	2211	CAGCCCAACGAGCGGCAAGCAGCTAGACCTGCTCTCTCGGTGACACTTACCCCAAGAC	2270
Db	1945	CAGGCCACAGCGCGCAAGCAGCTAGACCTGCTCTCTCGGTGACACTTACCCCAAGAC	2004
OY	2271	TGTGGACTGTGACTCATCTCCATCCGAGCGGTGGAGGTGGAGGTCTTACTGCTGTGCG	2330
Db	2005	TGTGGACTGTGACTCATCTCCATCCGAGCGGTGGAGGTGGAGGTCTTACTGCTGTGCG	2064
OY	2331	CCCTGGGCTCATCTTCTTGTGCTGTGTGAAAAAGAAAGAAAGAACAAACAGGCGCCCGC	2390
Db	2065	CCTGGGGCTCATCTTGTGCTGTGTGAAAAAGAAAGAAAGAACAAACAGGCGCCCGC	2124
OY	2391	TGTGGGTATCTCAATGCGCAATCAATACTGATGCCAGGACGCCAAAAAGTTTCA	2450
Db	2125	TGTGGGTATCTCAATGCGCAATCAATACTGATGCC - AGGACGCCAAAAAGTTTCA	2183
OY	2431	GAAGGGCGGAAGGACAATGACATCCCATGTGTGTGACATCTGAGGACACCATGTGTATA	2510
Db	2184	GAAGGGCGGAAGGACAATGACATCCCATGTGTGTGACATCTGAGGACACCATGTGTATA	2243
OY	2511	TGGGCAATCTGCTAACAGGATTTCCAGCGGCTCTTCTCGACGACAGAGTGGACACTACCG	2570
Db	2244	TGGGCAATCTGCTAACAGGATTTCCAGCGGCTCTTCTCGACGACAGAGTGGACACTACCG	2303
OY	2571	GCCGTTCCAGGGCACCATGGGGGGTCTGTCTCCCTCCCAACCACCATATGCTCCAGGGC	2630
Db	2304	GCCGTTCCAGGGCACCATGGGGGGTGTGTCTCTCCCTCCCAACCACCATATGCTCCAGGGC	2363
OY	2631	CCCAACTGCAAAAGTTGGCCCACTGAGAGAGCCACTCTGCGCTCCCTCTGAGTCTGAGAG	2690
Db	2364	CCCAACTGCAAAAGTTGGCCCACTGAGAGAGCCACTCTGCGCTCCCTCTGAGTCTGAGAG	2423
OY	2691	TGAACCGTACACCTTTCGCCATCCCAACAATGGGGATGTAAAGACAGAAGACACAGACAT	2750
Db	2424	TGAACCGTACACCTTTCGCCATCCCAACAATGGGGATGTAAAGACAGAAGACACAGACAT	2483
OY	2751	TCCCTTACTGAAACACTCAGAGAGCCCATGAGAGCCAGACAGATAACTGTGATTCATTCCAGAC	2810
Db	2484	TCCCTTACTGAAACACTCAGAGAGCCCATGAGAGCCAGACAGATAACTGTGATTCATTCCAGAC	2543
OY	2811	GCTTGGCTGAGTTTCATTAAGACAGGGGACTGAGACACCCGTCGGTGTCTTAACCAAGAA	2870
Db	2544	GCTTGGCTGAGTTTCATTAAGACAGGGGACTGAGACACCCGTCGGTGTCTTAACCAAGAA	2603
OY	2871	TCCCTAAGAAAGGAATTTATACAAAGGAAGACAGAGAGGTTTCCGTGGACACCGCCAAC	2930
Db	2604	TCCCTAAGAAAGGAATTTATACAAAGGAAGACAGAGAGGTTTCCGTGGACACCGCCAAC	2663
OY	2931	TTCACATTTGCTCAGTGGACTCATTTCTTAAGGGCAAGACATTTGAAAAATGATGAATTCCAATC	2990
Db	2664	TTCACATTTGCTCAGTGGACTCATTTCTTAAGGGCAAGACATTTGAAAAATGATGAATTCCAATC	2723
OY	2991	TGGATACAGTATATACACAGCTCATGTGCTCTCAACTTAAGGCTGTGCGGTTAGCCAGCTG	3050
Db	2724	TGGATACAGTATATACACAGCTCATGTGCTCTCAACTTAAGGCTGTGCGGTTAGCCAGCTG	2783
OY	3051	TAAAGAGAGGAGAGAGGCGTAGAGTCAACCTTGAATAGGTTGGAGCAAGCCCTGGAATTAG	3110
Db	2784	TAAAGAGAGGAGAGAGGCGTAGAGTCAACCTTGAATAGGTTGGAGCAAGCCCTGGAATTAG	2843
OY	3111	AGTGTTTAAACAGAGGCTTGGCCCTTTCAGGACAAAGTTTCCAATTTCCAAGAGGCTTCACT	3170
Db	2844	AGTGTTTAAACAGAGGCTTGGCCCTTTCAGGACAAAGTTTCCAATTTCCAAGAGGCTTCACT	2903
OY	3171	GAGGTCCCTACTCTCACTGGGGGTCCCAAGAGTGAACAACAATGTGCCCTTTTATATAT	3230
Db	2904	GAGGTCCCTACTCTCACTGGGGGTCCCAAGAGTGAACAACAATGTGCCCTTTTATATAT	2963
OY	3231	ATTATATTTGGGGGCGGTGTATTTTAAAGAGTCAAAATGTATAACACACTAGCTTTTTC	3290
Db	2964	ATTATATTTGGGGGCGGTGTATTTTAAAGAGTCAAAATGTATAACACACTAGCTTTTTC	3023
OY	3291	ACCTGACTAGTATATACATCATCTACTAACAGTGTTGGATCCCTGGGTTGTGACTTCTACTG	3350

Db	2700	GTATTAACACACTGACCTCTTTTCCACGACTTGTATATAATCAATCAATGTTGGAT	2759
Qy	3329	GCCGGGTTTGACTTCTACTGACCCGCTAGATAAAAGTGTCGCTTCCCCAGGGTGG	3388
Db	2760	GCCGGGTTTGACTTCTACTGACCCGCTAGATAAAAGTGTCGCTTCCCCAGGGTGG	2819
Qy	3389	GAATTAATTTACAAATCTGTCCACAGAAAAAGATGTGTGTTTGAGCAGCAATTGACA	3448
Db	2820	GAATTAATTTACAAATCTGTCCACAGAAAAAGATGTGTGTTTGAGCAGCAATTGACA	2879
Qy	3449	TATCTGTTTGATTAAGAGACTTCTGATTCCTAGTGGGTCTGTGGTTATCCATTGTG	3508
Db	2880	TATCTGTTTGATTAAGAGACTTCTGATTCCTAGTGGGTCTGTGGTTATCCATTGTG	2939
Qy	3509	GAATTCATCTTGAAATCCCAATGTGCTATAGTCCTGCAATTAAGAAATTTCCCAAGT	3568
Db	2940	GAATTCATCTTGAAATCCCAATGTGCTATAGTCCTGCAATTAAGAAATTTCCCAAGT	2999
Qy	3569	TTCCATGTGGGGTTCTCTAGCTGACAGCAATCTTGACATTTAAAGAAATTTAGAGA	3628
Db	3000	TTCCATGTGGGGTTCTCTAGCTGACAGCAATCTTGACATTTAAAGAAATTTAGAGA	3059
Qy	3629	ATATTCATCTCTTAAAAATGTTTAAATATATACCAACAGTGGCCCCCTGCATTAGT	3688
Db	3060	ATATTCATCTCTTAAAAATGTTTAAATATATACCAACAGTGGCCCCCTGCATTAGT	3119
Qy	3689	TTCTGTTGGCACTGCAACCCATTAATCTGGTAGCTTTAAAAACAACATTAAGCTTATAGT	3748
Db	3120	TTCTGTTGGCACTGCAACCCATTAATCTGGTAGCTTTAAAAACAACATTAAGCTTATAGT	3179
Qy	3749	CTGGGGATCAAAATTCGCAAAATGGATGTCCCTGAATGAATAATCAAGGTGCAGAGACT	3808
Db	3180	CTGGGGATCAAAATTCGCAAAATGGATGTCCCTGAATGAATAATCAAGGTGCAGAGACT	3239
Qy	3809	GTCGCTCTTGAAAGGCTCTAGGGAAACCCGGTTCCTGCCATTTCAAAGCTTCTAGAG	3868
Db	3240	GTCGCTCTTGAAAGGCTCTAGGGAAACCCGGTTCCTGCCATTTCAAAGCTTCTAGAG	3299
Qy	3869	CTGGCTCATTTCCAGGCTCACTGGCTGGTCAAGCTTTTCACATAGGACATCACTGGA	3928
Db	3300	CTGGCTCATTTCCAGGCTCACTGGCTGGTCAAGCTTTTCACATAGGACATCACTGGA	3359
Qy	3929	CACGTGGCCCTCCACTTCCTCTTTGACTTACAAAGCCACCAGGAAGATCCAGGATAT	3988
Db	3360	CACGTGGCCCTCCACTTCCTCTTTGACTTACAAAGCCACCAGGAAGATCCAGGATAT	3419
Qy	3989	CTCTCCATCTAAAGATCTCTTCAATCATCTGTGGAAGACCTTTTGCCATGCAAGCAACTA	4048
Db	3420	CTCTCCATCTAAAGATCTCTTCAATCATCTGTGGAAGACCTTTTGCCATGCAAGCAACTA	3479
Qy	4049	GCCACAGGTGGGATTTAGACACAGGACATCTTTGGGGTCTGTATTTGCTTACACAC	4108
Db	3480	GCCACAGGTGGGATTTAGACACAGGACATCTTTGGGGTCTGTATTTGCTTACACAC	3539
Qy	4109	CTTCTGCACTGACTCCACAGAGAGGCTCAAAATGATCTGGCGCACAGGATGTTT	4168
Db	3540	CTTCTGCACTGACTCCACAGAGAGGCTCAAAATGATCTGGCGCACAGGATGTTT	3599
Qy	4169	TGTTTAGCTTGGGAGCTCTAACTT - AAAAAAAAACCCGATCAGAAAGTCTGGCCATGC	4227
Db	3600	TGTTTAGCTTGGGAGCTCTAACTT AAAAAAAAACCCGATCAGAAAGTCTGGCCATGC	3659
Qy	4228	TGGGGCTACATCTCACCCTAGAACAAACACTGGCTGGAGTGGGGCACAGCTCTGCCCTTA	4287
Db	3660	TGGGGCTACATCTCACCCTAGAACAAACACTGGCTGGAGTGGGGCACAGCTCTGCCCTTA	3719
Qy	4288	GAAGGGGTGTCTCACTTACCAGGCTCAACAGGCCACATTAAGCCCTATCACTTCCACA	4347
Db	3720	GAAGGGGTGTCTCACTTACCAGGCTCAACAGGCCACATTAAGCCCTATCACTTCCACA	3779
Qy	4348	ATGAGGCTAAGTGTGTTTCTACTGATCAATGCCCTGCAGGTTGCATTTATTTGTAATG	4407

D	b	3780	ATGAGCCTAACTGTTTCTTCTCTCATCAATGAGCCCTGACAGTTGCATTATTTGTAATG	3839
Q	y	4408	AAAAAGAAAGACGTGGGATTTAATCTCAATCAGGTGAGTGGACCATAGAGACCAATGTGTGC	4467
D	b	3840	AAAAAGAAAGACGTGGATTTAATCTCTAATCAGGTGAGTGGACCATAGAGACCAATGTGTGC	3899
Q	y	4468	TCACATTACCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT	4527
D	b	3900	TCACATTACCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT	3959
Q	y	4528	GGATCTATTCTGTGTGCTGAGGCTGAGAGAGAGTGGGCGAATCTGGGCGACTGCAACT	4587
D	b	3960	GGATCTATTCTGTGTGCTGAGGCTGAGAGAGAGTGGGCGAATCTGGGCGACTGCAACT	4019
Q	y	4588	CTGGCTCTCTGGGCTCAAGCAATTTCTCCACCTGAGCTTCCAAATAGCTGGGATCACTG	4647
D	b	4020	CTGGCTCTCTGGGCTCAAGCAATTTCTCCACCTGAGCTTCCAAATAGCTGGGATCACTG	4079
Q	y	4648	CACAAACCAACAGCCAGCTAATTTTGTATTTTGTATGAGACAGGTTTCCACTT	4707
D	b	4080	CACAAACCAACAGCTAATTTTGTATTTTGTATGAGACAGGTTTCCACTT	4138
Q	y	4708	GCCGAGGCTGTCTACACGTCGAGGCTCAAGCAATCCGCGGCTCGGCGCCGCCCAAGT	4767
D	b	4139	GCCGAGGCTGTCTACACGTCGAGGCTCAAGCAATCCGCGGCTCGGCGCCGCCCAAGT	4198
Q	y	4768	GCTGGAGTACAGATGTGAGCCACCCGATCCAGCCCGACCCCTCATTTATACCAATTAC	4827
D	b	4199	GCTGGAGTACAGATGTGAGCCACCCGATCCAGCCCGACCCCGATTTATACCAATTAC	4258
Q	y	4828	CTGCCAGTAACTGTGAGCTTTTGTCTCTACCCCTGCTGTGATCTGGAGAGAGGGA	4887
D	b	4259	CTGCCAGTAACTGTGAGCTTTTGTCTCTACCCCTGCTGTGATCTGGAGAGAGGGA	4318
Q	y	4888	TTATGTATAGCTGTGAGAGAGCTCCAAAGTCAATTTCTGGGGGAAAACTCTCT	4947
D	b	4319	TTATGTATAGCTGTGAGAGAGCTCCAAAGTCAATTTCTGGGGGAAAACTCTCT	4378
Q	y	4948	CAAAAAATAAATGTAATCTTATTTCAATGAATTCACCTTGGAAATGACCCGCTCA	5007
D	b	4379	CAAAAAATAAATGTAATCTTATTTCAATGAATTCACCTTGGAAATGACCCGCTCA	4438
Q	y	5008	CTTGTTCACATGGCATTAATGAAAGAAATTTATAGTCTCTAATGGCGTGTACTGCA	5067
D	b	4439	CTTGTTCACATGGCATTAATGAAAGAAATTTATAGTCTCTAATGGCGTGTACTGCA	4498
Q	y	5068	GACGCTTGAACACTTTCCAGAGATATAGATATTAAAGTCAGCCCTGGGCTGCTCAT	5127
D	b	4499	GACGCTTGAACACTTTCCAGAGATATAGATATTAAAGTCAGCCCTGGGCTGCTCAT	4558
Q	y	5128	GGCACCTTCCCTTCAAAAGTCTGTGCTCTGGCCAGTACGCTTGGGCTTTGTGAGCGCA	5187
D	b	4559	GGCACCTTCCCTTCAAAAGTCTGTGCTCTGGCCAGTACGCTTGGGCTTTGTGAGCGCA	4618
Q	y	5188	GATGCTGACCTGCAATAAGGGCCAAAGAGGGGCTGGCTTCTTCCCTCACTGAAG	5247
D	b	4619	GATGCTGACCTGCAATAAGGGCCAAAGAGGGGCTGGCTTCTTCCCTCACTGAAG	4678
Q	y	5248	CCCTTATTTGAATCACTGTGTGTGAGGCGCTAGCCCTGCATTTCTCGACATTTCCCAACTC	5307
D	b	4679	CCCTTATTTGAATCACTGTGTGTGAGGCGCTAGCCCTGCATTTCTCGACATTTCCCAACTC	4738
Q	y	5308	CCAGCCCTTCAAGCAGAGCTAGTGCCCTGATTCACCCAAAGTGGGATTTGGCTTC	5367
D	b	4739	CCAGCCCTTCAAGCAGAGCTAGTGCCCTGATTCACCCAAAGTGGGATTTGGCTTC	4797
Q	y	5368	CTTAGGCTGGCTACTTTCTACACATCAACGACATCACTGTGTGCTGCAAGAGACACACG	5427
D	b	4798	CTTAGGCTGGCTACTTTCTACACATCAACGACATCACTGTGTGCTGCAAGAGACACACG	4857
Q	y	5428	GCCATTTTCTTCAACTGAGGGCTCAAAACTCTGGACAATTTGCTGGCTCTGAGACCA	5487
D	b	4858	GCCATTTTCTTCAACTGAGGGCTCAAAACTCTGGACAATTTGCTGGCTCTGAGACCA	4917

QY 5488 GTATTCCTGAGCTGTGGCTCACTGTAAGGGGCCAGCTGAGAACCTTGGCTCTTTC 5547
 Db 4918 GTATTCCTGAGCTGTGGCTCACTGTAAGGGGCCAGCTGAGAACCTTGGCTCTTTC 4977
 QY 5548 TTTAAGCCCGGCGCCCACTTACATTAACATTTTCAGGGTACTGGAACAGTAAGTC 5607
 Db 4978 TTTAAGCCCGGCGCCCACTTACATTAACATTTTCAGGGTACTGGAACAGTAAGTC 5037
 QY 5608 CATTTGTTGAAGCTGACTGATGCGAGCCCACTGCTCATCCAGTGGTCTGCATGCTTA 5667
 Db 5038 CATTTGTTGAAGCTGACTGATGCGAGCCCACTGCTCATCCAGTGGTCTGCATGCTTA 5097
 QY 5668 CGAGGAAGCCGAGCGCATGCGAGCACTGCTCTATGCTGTGTCATTGTCAGAGAGGA 5727
 Db 5098 CGAGGAAGCCGAGCGCATGCGAGCACTGCTCTATGCTGTGTCATTGTCAGAGAGGA 5157
 QY 5728 AAGGTCACAGGAAGATCACTGAGACAGACAGCCCAAGGACATGGCTTGGTAA 5787
 Db 5158 AAGGTCACAGGAAGATCACTGAGACAGACAGCCCAAGGACATGGCTTGGTAA 5217
 QY 5788 AGTTAGAGAGCTGCTGTGTGATGATGCGAGCTGCTCATCGAATTAATTATTCATTG 5847
 Db 5218 AGTTAGAGAGCTGCTGTGTGATGATGCGAGCTGCTCATCGAATTAATTATTCATTG 5277
 QY 5848 CAGATCTTTTAAAGTGGCATTTTATTCATTTCTGCTTTAAATAACAATGTACCA 5907
 Db 5278 CAGATCTTTTAAAGTGGCATTTTATTCATTTCTGCTTTAAATAACAATGTACCA 5337
 QY 5908 AAAACAAGATTCACAGCTTTTAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5967
 Db 5338 AAAACAAGATTCACAGCTTTTAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5397
 QY 5968 CCCGTTTCCAGTGTGTGATGATGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6027
 Db 5398 CCCGTTTCCAGTGTGTGATGATGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5457
 QY 6028 ATTTGGATGATACAGAAAGCCAGGCTTTGGATACAGTCTTCTCTTCAATTTGAT 6087
 Db 5458 ATTTGGATGATACAGAAAGCCAGGCTTTGGATACAGTCTTCTCTTCAATTTGAT 5517
 QY 6088 GCGGTGACGCTGTAAGAGCAGATGTTTTCGCCGAATAAATAATAGCTTGG 6143
 Db 5518 GCGGTGACGCTGTAAGAGCAGATGTTTTCGCCGAATAAATAATAGCTTGG 5573
 RESULT 5
 AC105902/c 163115 bp DNA linear PRI 27-JUN-2002
 LOCUS AC105902 Homo sapiens chromosome 3 clone RP11-578f5, complete sequence.
 DEFINITION AC105902
 ACCESSION AC105902.2 GI:21617714
 VERSION HTG.
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 163115)
 AUTHORS Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,
 Saenphimachak, C., Phelps, K.A., Buckley, D., Kibukawa, M., Raymond, C.
 and Haugen, E.D.
 TITLE Direct Submission
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 163115)
 AUTHORS Kaul, R.K., Olson, M.V., Raymond, C. and Haugen, E.D.
 TITLE Direct Submission
 JOURNAL Submitted (10-JAN-2002) Genome Center, University of Washington,
 Box 352145, Seattle, WA 98195, USA
 REFERENCE 3 (bases 1 to 163115)
 AUTHORS Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,
 Saenphimachak, C., Phelps, K.A., Buckley, D., Kibukawa, M., Raymond, C.
 and Haugen, E.D.
 TITLE Direct Submission

JOURNAL COMMENT

Submitted (27-JUN-2002) Genome Center, University of Washington,
 Box 352145, Seattle, WA 98195, USA
 On Jun 27, 2002 this sequence version replaced gi:18104827.
 ----- Genome Center
 Center: University of Washington Genome Center
 Center Code: UWGC
 Web site: <http://www.genome.washington.edu>
 Contact: uwgchgs@u.washington.edu
 ----- Project Information
 Center project name: chr-3
 Center clone name: RP11-578f5 (bc0477)
 ----- Summary Statistics
 Sequencing vector: plasmid; 100% of reads
 Chemistry: Dye-terminator ET; 2% of reads
 Chemistry: Dye-terminator Big Dye; 71% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 163063 bases at least Q40
 Consensus quality: 163115 bases at least Q30
 Insert size: 163115; sum-of-contigs
 Quality coverage: 9.1x in Q20 bases; sum-of-contigs

Overlapping Sequences:

5': RP11-68104 (UWGC:bc0528) AC104165
 3': RP11-487021 AC010170, 50065-bp overlap

Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:
 all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats: all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

ECORI

HindIII

BglII

SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt
-----	-----	-----	-----	-----	-----
8696	8823	3091	3097	3349	3352
-----	-----	-----	-----	-----	-----
6	<800	6382	6416	2067	2097
-----	-----	-----	-----	-----	-----
740	789	512	<800	6785	6921
-----	-----	-----	-----	-----	-----
10200	10022	449	<800	50	<800
-----	-----	-----	-----	-----	-----
1644	1614	5235	5065	1052	1030
-----	-----	-----	-----	-----	-----
9393	9315	2845	2873	852	842
-----	-----	-----	-----	-----	-----

[illegible]

QY 2946 GGAATCTCTTAAGGGGCAAGACATTTGAAATGATGAATTCATCTGGATACAGTCATGA 3005
 |||||
 Db 73621 GGAATCTCTTAAGGGGCAAGACATTTGAAATGATGAATTCATCTGGATACAGTCATGA 73562
 QY 3006 CAGCTCATGTGCTCTTAACCTTAGGCTGTGGCTGACCCAGCCCTGTAATAGAGAGAGA 3065
 |||||
 Db 73561 CAGCTCATGTGCTCTTAACCTTAGGCTGTGGCTGACCCAGCCCTGTAATAGAGAGAGA 73502
 QY 3066 GGCCTGAGTCACTAGCATGAGGTTGACAGAAAGCCCTGATTCAGAGTGTAAACAGAGG 3125
 |||||
 Db 73501 GGCCTGAGTCACTAGCATGAGGTTGACAGAAAGCCCTGATTCAGAGTGTAAACAGAGG 73442
 QY 3126 CTGGCCCTCTTCAAGGACACAGTTCCAAATTCAGAGAGCCCTACCTGAGGTCCTACTCTC 3185
 |||||
 Db 73441 CTGGCCCTCTTCAAGGACACAGTTCCAAATTCAGAGAGCCCTACCTGAGGTCCTACTCTC 73382
 QY 3186 ACAGGGGTCCTCCAGAGATGAAAGACAAATGTCCTTTTATATATATATTTGTTGTTGTC 3245
 |||||
 Db 73381 ACAGGGGTCCTCCAGAGATGAAAGACAAATGTCCTTTTATATATATATTTGTTGTTGTC 73322
 QY 3246 CTGTGTATTTAAGAGATCAAAATGTATACACCTAGCTCTTTTACCTGACCTAGTAAT 3305
 |||||
 Db 73321 CTGTGTATTTAAGAGATCAAAATGTATACACCTAGCTCTTTTACCTGACCTAGTAAT 73262
 QY 3306 AACTCATACTACTGCTGTTGGATGCTGGGTTGTGACTTCTACTGACCGCTAGATAACG 3365
 |||||
 Db 73261 AACTCATACTACTGCTGTTGGATGCTGGGTTGTGACTTCTACTGACCGCTAGATAACG 73202
 QY 3366 TGAGCCCTGCTCCAGAGTGGGATTAATTTCAATCTGTCACCAAGCAAGAAAGAAAGTG 3425
 |||||
 Db 73201 TGAGCCCTGCTCCAGAGTGGGATTAATTTCAATCTGTCACCAAGCAAGAAAGAAAGTG 73142
 QY 3426 TGTGTTTGAAGAGATGACACATATCTGCTTGTGATTAAGAGACTTCCGTGATTTCTAGT 3485
 |||||
 Db 73141 TGTGTTTGAAGAGATGACACATATCTGCTTGTGATTAAGAGACTTCCGTGATTTCTAGT 73082
 QY 3486 CGGTCTGCTGTTATCCCATTTGTGAAATTCATCTTGAATCCCATTTGCTATAGTCTTAG 3545
 |||||
 Db 73081 CGGTCTGCTGTTATCCCATTTGTGAAATTCATCTTGAATCCCATTTGCTATAGTCTTAG 73022
 QY 3546 CAATTAAGAGAAATTTCTCAAGTTTCATGAGGTTCTCTTACCTGACAGCAATACTTG 3605
 |||||
 Db 73021 CAATTAAGAGAAATTTCTCAAGTTTCATGAGGTTCTCTTACCTGACAGCAATACTTG 72962
 QY 3606 ACATTTAAGAGAAATTTAAGAGATATTTCTATCTCTAATAAATGTTTAAATATATACCA 3665
 |||||
 Db 72961 ACATTTAAGAGAAATTTAAGAGATATTTCTATCTCTAATAAATGTTTAAATATATACCA 72902
 QY 3666 AACAGTGGCCCTGCAATTAAGTTTCTGTTGCCACTGCAACCCATTTCTTGTAAGCTTAA 3725
 |||||
 Db 72901 AACAGTGGCCCTGCAATTAAGTTTCTGTTGCCACTGCAACCCATTTCTTGTAAGCTTAA 72842
 QY 3726 AAACACACATTAAGCTTATAGTCTCTGGGATCAGAAATTCACAAATGATGCCGTGATG 3785
 |||||
 Db 72841 AAACACACATTAAGCTTATAGTCTCTGGGATCAGAAATTCACAAATGATGCCGTGATG 72782
 QY 3786 AAAATCAAGGTGTCAGAGAGCTGTGCTCTTCTGAAGGCTTAGGAGAGAGCCGGTTCC 3845
 |||||
 Db 72781 AAAATCAAGGTGTCAGAGAGCTGTGCTCTTCTGAAGGCTTAGGAGAGAGCCGGTTCC 72722
 QY 3846 TTGGCATTTCAAGCTTCTAGAGGCTGGCTGCAATTCACAGGCTCAGTGGCTCAAGCT 3905
 |||||
 Db 72721 TTGGCATTTCAAGCTTCTAGAGGCTGGCTGCAATTCACAGGCTCAGTGGCTCAAGCT 72662
 QY 3906 TTTTCTCATGGGATCCTACTGTGACACTGAGCCCTCCCACTTCCCTTTGACTTGAATCAAGC 3965
 |||||
 Db 72661 TTTTCTCATGGGATCCTACTGTGACACTGAGCCCTCCCACTTCCCTTTGACTTGAATCAAGC 72602
 QY 3966 CCACACAGAGATCCAGAGATTAATCTCTCAATCTTAAGATCCTTATCATCTGGAAGAGC 4025
 |||||
 Db 72601 CCACACAGAGATCCAGAGATTAATCTCTCAATCTTAAGATCCTTATCATCTGGAAGAGC 72542

QY 4026 CTTTGGCATGCAAGACACATTAACCAAGGTGGGATTAAGGACAGAGACATCTTTGGG 4085
 |||||
 Db 72541 CTTTGGCATGCAAGACACATTAACCAAGGTGGGATTAAGGACAGAGACATCTTTGGG 72482
 QY 4086 TGGCTTATTTCTGCTTACCAACACCTTCTGCACTGTGCTCCACAGAGAGGCTACAAA 4145
 |||||
 Db 72481 TGGCTTATTTCTGCTTACCAACACCTTCTGCACTGTGCTCCACAGAGAGGCTACAAA 72422
 QY 4146 TGAATCTGGCCACAGAGGATTTTGTATTTAGCTTGGGACTCTAACCTT-AAAAAACC 4204
 |||||
 Db 72421 TGAATCTGGCCACAGAGGATTTTGTATTTAGCTTGGGACTCTAACCTTAAAAAACC 72362
 QY 4205 CAGATCAGAAATCTGGCCATGCTGGGCTCACATTTCTACATAGCAACAACCTGGCTGA 4264
 |||||
 Db 72361 CAGATCAGAAATCTGGCCATGCTGGGCTCACATTTCTACATAGCAACAACCTGGCTGA 72302
 QY 4265 GCTGGGACACAGCTCTGCTTTAGAAAGGGGTGTCATTCACAGGTCACACAGCCAC 4324
 |||||
 Db 72301 GCTGGGACACAGCTCTGCTTTAGAAAGGGGTGTCATTCACAGGTCACACAGCCAC 72242
 QY 4325 ACTAGCCCTATCACTTCCCAATGAGGCTAAGTGTGTTCTACTGATCAATGCCCC 4384
 |||||
 Db 72241 ACTAGCCCTATCACTTCCCAATGAGGCTAAGTGTGTTCTACTGATCAATGCCCC 72182
 QY 4385 TGCAGTTGCATTTATTTGTAATGAAAAAGAAAGACTGGATTAATCTTAATCAGGTGAG 4444
 |||||
 Db 72181 TGCAGTTGCATTTATTTGTAATGAAAAAGAAAGACTGGATTAATCTTAATCAGGTGAG 72122
 QY 4445 TAGACATGAGACCAATGTGCTCACATTAACCTTTTCTTTTCTTTTCTTTTCTTTT 4504
 |||||
 Db 72121 TAGACATGAGACCAATGTGCTCACATTAACCTTTTCTTTTCTTTTCTTTTCTTTT 72062
 QY 4505 CTTTTTTTTTAAATGTAAGACAGATCTATCTGTTGCTAGGCTGAGATGAGTGCC 4564
 |||||
 Db 72061 CTTTTTTTTTAAATGTAAGACAGATCTATCTGTTGCTAGGCTGAGATGAGTGCC 72002
 QY 4565 GCAATCTGGCTCACTGCAACCTCTGCTCTGGGCTCAAGCAATTTCTCCACTCAGCC 4624
 |||||
 Db 72001 GCAATCTGGCTCACTGCAACCTCTGCTCTGGGCTCAAGCAATTTCTCCACTCAGCC 71942
 QY 4625 TCCCAATTAAGTGGGATCACTGCAACCAACCAACCACTCCAGTAATTTTGTATTTTGG 4684
 |||||
 Db 71941 TCCCAATTAAGTGGGATCACTGCAACCAACCAACCACTCCAGTAATTTTGTATTTTGG 71882
 QY 4685 TAGAGACAGGTTTACCATGTTGCCAGGCTGTCTCAACCTCTGSGCTCAAGCAATC 4744
 |||||
 Db 71881 TAGAGACAGGTTTACCATGTTGCCAGGCTGTCTCAACCTCTGSGCTCAAGCAATC 71822
 QY 4745 CTCTGGCTGGCTCCCAAGGCTGGGATTAACAGATGAGGACAGGACAGGCCCC 4804
 |||||
 Db 71821 CTCTGGCTGGCTCCCAAGGCTGGGATTAACAGATGAGGACAGGACAGGCCCC 71762
 QY 4805 ACACCCCTATTTATACCAATTAACCTGCCAGTAACCTGAGCTTCTCTTCAACCCCT 4864
 |||||
 Db 71761 ACACCCCTATTTATACCAATTAACCTGCCAGTAACCTGAGCTTCTCTTCAACCCCT 71702
 QY 4865 GCTCTGATCTGGAAGAGAGGATTAATTAATGATCTTGTGACGACAGCTGCCAAGTCAAT 4924
 |||||
 Db 71701 GCTCTGATCTGGAAGAGAGGATTAATTAATGATCTTGTGACGACAGCTGCCAAGTCAAT 71642
 QY 4925 ATTTCTGGGCAAAAATCTTCTTCAAAAATTAATTAATGACTTCAATTTGTAATGAATTC 4984
 |||||
 Db 71641 ATTTCTGGGCAAAAATCTTCTTCAAAAATTAATTAATGACTTCAATTTGTAATGAATTC 71582
 QY 4985 ACCTTGAATTAAGACCCCTCAACTTTTTCACATGAGATTAATGAAGAAATTTATAGT 5044
 |||||
 Db 71581 ACCTTGAATTAAGACCCCTCAACTTTTTCACATGAGATTAATGAAGAAATTTATAGT 71522
 QY 5045 CTCTTAATTAAGGCTGTACTGCAAGACCTCTTGAACATTTTCAAGAGATAGGATATTTAA 5104
 |||||
 Db 71521 CTCTTAATTAAGGCTGTACTGCAAGACCTCTTGAACATTTTCAAGAGATAGGATATTTAA 71462
 QY 5105 GTCTAGCCCTTGGGCTTGCCTATGACCTTTTCCCTTCAAAAGTGTGTTCTGCCACG 5164

Db	71461	GTCAATGCCCTTGGCGTGGCTATGGACACTTTCCTTGTGAAAGTCGTGGTCTCCGCCAG	714020
OY	5165	TGACCCCTTGGCCTTGTGAGCCGAGATGCTGACCTGTGACATAAAGGCCAAAGAGGGCTGCG	5224
Db	71401	TGACCCCTTGGCCTTGTGAGCCGAGATGCTGACCTGTGACATAAAGGCCAAAGAGGGCTGCG	713424
OY	5225	GGCTTCCTTCCCTGACTGAAGAGCCCTTATTGATATTCACGTGTGGAGCCCTAACCCCTC	5284
Db	71341	GGCTTCCTTCCCTGACTGAAGAGCCCTTATTGATATTCACGTGTGGAGCCCTAACCCCTC	712822
OY	5285	CATTCTCGACATYTCGCCAACCTCCAGGCCCTTCCACAGCAGACTAGAGTCCCTGCATTC	5344
Db	71281	CATTCTCGACATYTCGCCAACCTCCAGGCCCTTCCACAGCAGACTAGAGTCCCTGCATTC	712222
OY	5345	CACCCAGGTGGGATYTGSCCTTCTTTAGGCTGGCTACTTGTCCACATCACCGACATCACT	5404
Db	71221	CACCCAGGTGGGATYTGSCCTTCTTTAGGCTGGCTACTTGTCCACATCACCGACATCACT	711622
OY	5405	GTTGGCTCAGAGGACACACCGTGGCCATTTTCCCTCACTGAGGGCTCAAAACTCCTGGA	5464
Db	71161	GTTGGCTCAGAGGACACACCGTGGCCATTTTCCCTCACTGAGGGCTCAAAACTCCTGGA	711020
OY	5465	CAAGTTCGTGGCTCCTGAGACAGATATTTCCGAGAGCTGTGGCTAGTGAAGGGGCCAG	5524
Db	71101	CAAGTTCGTGGCTCCTGAGACAGATATTTCCGAGAGCTGTGGCTAGTGAAGGGGCCAG	710422
OY	5525	CTTGAGGAACCCCTGGCTTTTCTTTTAAAGCCAGGCCCCCACTTACATAAACATTTTCAG	5584
Db	71041	CTTGAGGAACCCCTGGCTTTTCTTTTAAAGCCAGGCCCCCACTTACATAAACATTTTCAG	709822
OY	5585	GGTCACTGGGAAACAGTGAAGTGCATTTTGTGTGAACCTTACTGATGCCAGCCACTGCTC	5644
Db	70981	GGTCACTGGGAAACAGTGAAGTGCATTTTGTGTGAACCTTACTGATGCCAGCCACTGCTC	709222
OY	5645	ATCCACGTGTGCTGCGCATAGCCTACAGAGAAGGCCAGCCATGCAGAGCTGGCTCTAATG	5704
Db	70921	ATCCACGTGTGCTGCGCATAGCCTACAGAGAAGGCCAGCCATGCAGAGCTGGCTCTAATG	708622
OY	5705	CTGTGCTATTGTCACAGAGAAGGAAGTCTCAAGGAAGTCTACCTGGGACAAGCACAAG	5764
Db	70861	CTGTGCTATTGTCACAGAGAAGGAAGTCTCAAGGAAGTCTACCTGGGACAAGCACAAG	708022
OY	5765	CCGACCCGACATGGGCTTGTGAAGGTGAAGCACTGCTGTGTGTGGATCTGCACTGCTT	5824
Db	70801	CCGACCCGACATGGGCTTGTGAAGGTGAAGCACTGCTGTGTGTGGATCTGCACTGCTT	707422
OY	5825	CACGTGAATTAATTATTTCATGTGACGATACCTTTTAGTGGGCACTTTATTCTATTCCTGT	5884
Db	70741	CACGTGAATTAATTATTTCATGTGACGATACCTTTTAGTGGGCACTTTATTCTATTCCTGT	706822
OY	5885	GCTTTAAATTAACAATGTACCAAAAAACAAGTATCAAGCTGTTTAAAGTCTTGGCTTAC	5944
Db	70681	GCTTTAAATTAACAATGTACCAAAAAACAAGTATCAAGCTGTTTAAAGTCTTGGCTTAC	706222
OY	5945	TTTGTCCTCCCTGGTTCAGTGAAGGCCGCCGGTTTCCAGTTGTGTGACTGTGACAGGCTCAGCA	6004
Db	70621	TTTGTCCTCCCTGGTTCAGTGAAGGCCGCCGGTTTCCAGTTGTGTGACTGTGACAGGCTCAGCA	705622
OY	6005	TGGGCTCAGCAGATGCTGTCTTAATTGTGTGATGTATACAGAAAGCAGAGCTTTGGGATAC	6064
Db	70561	TGGGCTCAGCAGATGCTGTCTTAATTGTGTGATGTATACAGAAAGCAGAGCTTTGGGATAC	705022
OY	6065	AAGTTCCTTCCCTCTTCATTTGATGCGCGTGAAGTGTGTGAAGCAGATGTTTTGTCCGGA	6124
Db	70501	AAGTTCCTTCCCTCTTCATTTGATGCGCGTGAAGTGTGTGAAGCAGATGTTTTGTCCGGA	704422
OY	6125	ATAAAATAATATGTCCTTGGAGTCTTCGCCA 6153	
Db	70441	ATAAAATAATATGTCCTTGGAGTCTTCGCCA 70413	

RESULT 6

LOCUS	AC104165/c	200956 bp	DNA	linear	PRI 27-JUN-2002
DEFINITION	Homo sapiens chromosome 3 clone RP11-68104, complete sequence.				
ACCESSION	AC104165 AC068625				
VERSION	AC104165.2 GI:21617715				
KEYWORDS	HTG.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	1 (bases 1 to 200956) Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z., Saenphlammachak, C., Phelps, K.A., Buckley, D., Kibukawa, M., Raymond, C. and Haugen, E.D.				
TITLE	Direct Submmission				
JOURNAL	2 (bases 1 to 200956) unpublished				
REFERENCE	Kaul, R.K., Olson, M.V., Raymond, C. and Haugen, E.D.				
AUTHORS	Direct Submmission				
TITLE	Submitted (03-DEC-2001) Genome Center, University of Washington,				
JOURNAL	Box 352145, Seattle, WA 98195, USA				
REFERENCE	3 (bases 1 to 200956) Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z., Saenphlammachak, C., Phelps, K.A., Buckley, D., Kibukawa, M., Raymond, C. and Haugen, E.D.				
AUTHORS	Direct Submmission				
TITLE	Submitted (27-JUN-2002) Genome Center, University of Washington,				
JOURNAL	Box 352145, Seattle, WA 98195, USA				
REFERENCE	On Jun 27, 2002 this sequence version replaced gi:17352429.				
COMMENT	----- Genome Center				
TITLE	Center: University of Washington Genome Center				
JOURNAL	Center Code: UWGC				
REFERENCE	Web site: http://www.genome.washington.edu				
AUTHORS	Contact: uwgchgs@u.washington.edu				
TITLE	Drafting Center: WUGSC				
JOURNAL	----- Project Information				
REFERENCE	Center project name: Chr-3				
AUTHORS	Center clone name: RP11-68104 (bc0528)				
TITLE	----- Summary Statistics				
JOURNAL	Sequencing vector: unknown; 42% of reads				
REFERENCE	Sequencing vector: plasmid; 58% of reads				
AUTHORS	Chemistry: Dye-terminator ET; 80% of reads				
TITLE	Chemistry: Dye-terminator Big Dye; 20% of reads				
JOURNAL	Assembly program: Phrap; version 0.990319				
REFERENCE	Consensus quality: 200732 bases at least Q40				
AUTHORS	Consensus quality: 200909 bases at least Q30				
TITLE	Consensus quality: 200947 bases at least Q20				
JOURNAL	Insert size: 200956; sum-of-contigs				
REFERENCE	Quality coverage: 10.4x in Q20 bases; sum-of-contigs				
AUTHORS	-----				
TITLE	Overlapping Sequences:				
JOURNAL	5': RP11-348P10 (UWGC:bc0377) AC124045				
REFERENCE	3': RP11-578F5 (UWGC:bc0477) AC105902, 95352-bp overlap				
AUTHORS	-----				
TITLE	Sequence Quality Assessment:				
JOURNAL	This entry has been annotated with sequence quality				
REFERENCE	estimates computed by the Phrap assembly program.				
AUTHORS	All manually edited bases have been reduced to quality zero.				
TITLE	Quality levels above 40 are expected to have less than				
JOURNAL	1 error in 10,000 bp.				
REFERENCE	Base-by-base quality values are not generally visible from the				
AUTHORS	Genbank flat file format but are available as part				
TITLE	of this entry's ASN.1 file.				
JOURNAL	-----				
REFERENCE	This sequence was finished as follows unless otherwise noted:				
AUTHORS	all regions were either double-stranded or sequenced with an				
TITLE	alternate chemistry or covered by high quality data (i.e., Phred				
JOURNAL	quality >= 30); an attempt was made to resolve all sequencing				
REFERENCE	problems, such as compressions and repeats: all regions were				
AUTHORS	covered by at least one plasmid subclone or more than one M13				
TITLE	subclone; and the assembly was confirmed by restriction digest.				
JOURNAL	-----				
REFERENCE	Sequence Validation:				
AUTHORS	-----				

D	179746	GGATTCCAGCGGCTCTTCTCTCGACGCCAGAGTGGACACTACCGGCGCTTCACGGGCAC	179687
Q	2386	CATGGGGGTCTGTCTCTCCCTCCACCACCACCATATGCTCCAGGGGCCCACTGCAAGTT	2645
D	179686	CATGGGGGTCTGTCTCTCCCTCCACCACCACCATATGCTCCAGGGGCCCACTGCAAGTT	179627
Q	2646	GGCCACTGAGGACGCCCTCTCGCTCTCCCTGAGCTGAGAGTGAACCGTACACTT	2705
D	179626	GGCCACTGAGGAGCCACTCTCTCGCTCTCCCTGAGCTGAGAGTGAACCGTACACTT	179567
Q	2706	CTCCCATCCCAAAATGGGGGTGTAACGACGACGACACGACATTCCTCTACTGAAAC	2765
D	179566	CTCCCATCCCAAAATGGGGGTGTAACGACGACGACACGACATTCCTCTACTGAAAC	179507
Q	2766	TCAGAGGCCCATGAGAGCCAGCAGAAATAACTGTGATCAATTCACAGCGCTTGTCTGAGTTTC	2825
D	179506	TCAGAGGCCCATGAGAGCCAGCAGAAATAACTGTGATCAATTCACAGCGCTTGTCTGAGTTTC	179447
Q	2826	ATAAAGCAGGGCCTGAGACACCCGTCGCTGTTCTTAACGAAATCTTAAGAAGAGA	2885
D	179446	ATAAAGCAGGGCCTGAGACACCCGTCGCTGTTCTTAACGAAATCTTAAGAAGAGA	179387
Q	2886	ATTATACAGAAAGAAACGAGGAGGTTTCTTGACACCGCCCACTTCACATTGCTCAGT	2945
D	179386	ATTATACAGAAAGAAACGAGGAGGTTTCTTGACACCGCCCACTTCACATTGCTCAGT	179327
Q	2946	GGACTCATTTCTTAAGGGCAGACATTTGAAATGATTAATTCCAATCTGGATACAGTCTATGA	3005
D	179326	GGACTCATTTCTTAAGGGCAGACATTTGAAATGATTAATTCCAATCTGGATACAGTCTATGA	179267
Q	3006	CAGCTCATGTGCTCTCTCAACTTAGGCTGTGCGGTTAGCCAGCCGTGAATGAGAGGAGA	3065
D	179266	CAGCTCATGTGCTCTCTCAACTTAGGCTGTGCGGTTAGCCAGCCGTGAATGAGAGGAGA	179207
Q	3066	GGCGTGAAGTCAACCGATAGAGGTTGGACGACAGCCCTGAGATTCAGAGTGTAAACGAGG	3125
D	179206	GGCGTGAAGTCAACCGATAGAGGTTGGACGACAGCCCTGAGATTCAGAGTGTAAACGAGG	179147
Q	3126	CTTGCCCGCTTTCAGAGACAACAGTTTCCAAATTCACAAGAGACCTACCTGAGGCGCCCTACTCTC	3185
D	179146	CTTGCCCGCTTTCAGAGACAACAGTTTCCAAATTCACAAGAGACCTACTGAGGCGCCCTACTCTC	179087
Q	3186	ACTGGGGTCCCCAGAGTAAAAACGACAAATGTGCTTTTATTATTATTATTGGTGGTC	3245
D	179086	ACTGGGGTCCCCAGAGTAAAAACGACAAATGTGCTTTTATTATTATTATTGGTGGTC	179027
Q	3246	CTGTGTTATTAAAGATCAAAATGTATATACCACTACTCTTTTACCTGACTTATAT	3305
D	179026	CTGTGTTATTAAAGATCAAAATGTATATACCACTACTCTTTTACCTGACTTATAT	179677
Q	3306	AACATCATCTAACTGTTTGGATGCTCTGGGTTGTGACTTCTACTGACCGGTAGATAAGC	3365
D	178966	AACATCATCTAACTGTTTGGATGCTCTGGGTTGTGACTTCTACTGACCGGTAGATAAGC	178907
Q	3366	TGTGCTGTGCTCCCGCAGGTGTGTGGAAATTAATTACAAATCTGTCCAAACGAAAAAATGTG	3425
D	178906	TGTGCTGTGCTCCCGCAGGTGTGTGGAAATTAATTACAAATCTGTCCAAACGAAAAAATGTG	178847
Q	3426	TGTGTTTGAGAGACGATTGACACATATCTGCTTTGATTAAGAGACTTCTGTATTTCTTAAGT	3485
D	178846	TGTGTTTGAGAGACGATTGACACATATCTGCTTTGATTAAGAGACTTCTGTATTTCTTAAGT	178787
Q	3486	CGGTTGCGGGTTATCCCATTTGTGGAATTCATCTTGAATCCCATGTGCTATATGTCTAG	3545
D	178786	CGGTTGCGGGTTATCCCATTTGTGGAATTCATCTTGAATCCCATGTGCTATATGTCTAG	178727
Q	3546	CAATTAAGAAATTTCCCTCAAGTTTCCATGTGGCGTTCTCTAGCTGACACCAATTAATTG	3605
D	178726	CAATTAAGAAATTTCCCTCAAGTTTCCATGTGGCGTTCTCTAGCTGACACCAATTAATTG	178667
Q	3606	ACATTTAAGAAATTTAGAGATTAATTCATCTCTTAAAAATGTTTAAATATATACCA	3665

Db 178666	ACATTTAAAGAGAAATTTAGAGAAATATCTCATCTCTCTATAAAATGTTTAAATATATACCA	178607
QY	3666 AACAGTGGCCCCCTGCATTAAGTTTCTGTGGCCACGTGCAACCCATTACTTGGTAGACTTAA	3725
Db 178606	AACAGTGGCCCCCTGCATTAAGTTTCTGTGGCCACGTGCAACCTATTACTTGGTAGACTTAA	178547
QY	3726 AAACAACACATTAGCTTATATAGTCTCTGGGGATCGAAATTCAGAAATGGAATGTCTCTGAATG	3785
Db 178546	AAACAACACATTAGCTTATATAGTCTCTGGGGATCGAAATTCAGAAATGGAATGTCTCTGAATG	178487
QY	3786 AAAATTAAGGTGTACACAGAGCTGTGGTCCCTGTGAAAGGCTTAGGGAGAAAGCCGGTTC	3845
Db 178486	AAAATTAAGGTGTACACAGAGCTGTGGTCCCTGTGAAAGGCTCTTAGGGAGAAAGCCGGTTC	178427
QY	3846 TTGGCATTTTCAACCTTCTAGAGGCTGGCTGCATTTCCAGGCTCCAGTGGCTGTCAAAGCT	3905
Db 178426	TTGGCATTTTCAACCTTCTAGAGGCTGGCTGCATTTCCAGGCTCCAGTGGCTGTCAAAGCT	178367
QY	3906 TTTTCTCACATGGCATCAGTGTGACACTGGCCCTCCCACTTCCCTCTTTGACTTTACAAAGC	3965
Db 178366	TTTTCTCACATGGCATCAGTGTGACACTGGCCCTCCCACTTCCCTCTTTGACTTTACAAAGC	178307
QY	3966 CCACACGAGGAATCCAGGGAATATCTGCCATCTGAAAGATCCTTCATCATCTCGGAAAGC	4025
Db 178306	CCACACGAGGAATCCAGGGAATATCTGCCATCTGAAAGATCCTTCATCATCTCGGAAAGC	178247
QY	4026 CTTTTCACATGCAGAGCAACATATAGCCACAGTGGGGATTAGAGACAGACATCTTTGGG	4085
Db 178246	CTTTTCACATGCAGAGCAACATATAGCCACAGTGGGGATTAGAGACAGACATCTTTGGG	178187
QY	4086 TGCCTGTTATTTCTGCTTACACACACCTTCTGCCACTGTACTCCACAGAGAGGCTTACAAA	4145
Db 178186	TGCCTGTTATTTCTGCTTACACACACCTTCTGCCACTGTACTCCACAGAGAGGCTTACAAA	178127
QY	4146 TGAATCGGGGACACAGGGAATGTTTGTATACCTGGGACCTTAACACTT-AAAAAAACC	4204
Db 178126	TGAATCGGGGACACAGGGAATGTTTGTATACCTGGGACCTTAACACTTAAAAAAAACC	178067
QY	4205 CAGATCAGAGAATCTGGGCATGCTGGGGCTCACATTTCTACCTAGCAACAACTGGCTGGA	4264
Db 178066	CAGATCAGAGAATCTGGGCATGCTGGGGCTCACATTTCTACCTAGCAACAACTGGCTGGA	178007
QY	4265 GCTGGGACACAGCTGTGCTTTAGAAAGGGGTGTCCACTTACACAGGTCCACACAGCCAC	4324
Db 178006	GCTGGGACACAGCTGTGCTTTAGAAAGGGGTGTCCACTTACACAGGTCCACACAGCCAC	177947
QY	4325 ACTAGCCCCATATCACTTCCCAATAGGCTAGTGTCTTTCTACTGATCAATGGCCC	4384
Db 177946	ACTAGCCCCATATCACTTCCCAATAGGCTAGTGTCTTTCTACTGATCAATGGCCC	177887
QY	4385 TGCAGGTGCATTTATTTGTAATGAAAGAAAGAAACGTGGATTAATCTCTAATCAGTGAG	4444
Db 177886	TGCAGGTGCATTTATTTGTAATGAAAGAAAGAAACGTGGATTAATCTCTAATCAGTGAG	177827
QY	4445 TAGACATGAGACCAATGTGTGCTCACAATTACCCCTTTTCTTTTTTCTTTTTCTTTTT	4504
Db 177826	TAGACATGAGACCAATGTGTGCTCACAATTACCCCTTTTCTTTTTTCTTTTTCTTTTT	177767
QY	4505 CTTTTTTTTTTTTAATGTAGACAGAGATCTATTTCTGTGGCTTAGGCTGAGGTGAGTGGC	4564
Db 177766	CTTTTTTTTTTTTTAATGTAGACAGAGATCTATTTCTGTGGCTTAGGCTGAGGTGAGTGGC	177707
QY	4565 GCAATCTGGGCTACGTGCAACCTGTGCTCTGGGGCTCAAGCAATTCTCCACCTCAGCC	4624
Db 177706	GCAATCTGGGCTACGTGCAACCTGTGCTCTGGGGCTCAAGCAATTCTCCACCTCAGCC	177647
QY	4625 TCCCAAAATAGCTGGATCACTGGCACAAACACACATGCCCCAGCTAATTTTGTATTTTTTG	4684
Db 177646	TCCCAAAATAGCTGGATCACTGGCACAAACACACATGCCCCAGCTAATTTTGTATTTTTTG	177587
QY	4685 TAGAGACAGGTTTCAACCATGTTGCCAGGCTGTCTCAACCTCTCTGGGCTCAAGCATATC	4744
Db 177586	TAGAGACAGGTTTCAACCATGTTGCCAGGCTGTCTCAACCTCTCTGGGCTCAAGCATATC	177527

OY	4745	CTCTGCTCGGCGCTCCCAAGTCTGGGATTAAGATGTAGCACCGATCCAGCCC	4804
OY	4745	CTCTGCTCGGCGCTCCCAAGTCTGGGATTAAGATGTAGCACCGATCCAGCCC	4804
Db	177526	CTCCGCTCGGCGCTCCCAAGTCTGGGATTAAGATGTAGCACCGATCCAGCCC	177467
OY	4805	ACACCCTATTATACCAATTACCTGCCAGTAAGTGGACCTTTCTCTCCACCCT	4864
Db	177466	ACACCCTATTATATACCAATTACCTGCCAGTAAGTGGACCTTTCTCTCCACCCT	177407
OY	4865	GCTCGATCTGGGAAGGAGGATTAAGTTATACCTGTACGACACTCCCAACTTCAT	4924
Db	177406	GCTCGATCTGGGAAGGAGGATTAAGTTATACCTGTCTGACACACTCCCAACTTCAT	177347
OY	4925	ATTCTGGGGCAAAACACTCTCTCAAAAAATAATGTACTCATGTATTCAAATATC	4984
Db	177346	ATTCTGGGGCAAAACACTCTCTCAAAAAATAATGTACTCATGTATTCAAATATC	177287
OY	4985	ACCTTGGAAATGCACCGCTCTCACTGTTCACATGCAATTAATGAAGATTTATAGT	5044
Db	177286	ACCTTGGAAATGCACCGCTCTCACTGTTCACATGCGATTAATGAAGATTTATAGT	177227
OY	5045	CTCCCAATGGCGGTACTGTGAAGCCCTTGAACACTTTCACAGAGATATGATATTA	5104
Db	177226	CTCCCAATGGCGGTACTGTGAAGCCCTTGAACACTTTCACAGAGATATGATATTA	177167
OY	5105	GTCATGCCCTTGGGCTTCCCTATGGACCTTCCCTCTGAAGTCTGGTCTCGCCAG	5164
Db	177166	GTCATGCCCTTGGGCTTCCCTATGGACCTTCCCTCTGGAAGTCTGGTCTCGCCAG	177107
OY	5165	TGACCCCTGGGCTGTGAGCCGCAATGCTGACCTCGCATAAAGGCCAAAGAGGCTGC	5224
Db	177106	TGACCCCTGGGCTGTGAGCCGCAATGCTGACCTCGCATAAAGGGCCAAAGAGGCTGC	177047
OY	5225	GGCCTTCTTCCCTCACTAAGAGCCCTTATTTGAATTCACTGTGTGAGCCCTAGCCCTC	5284
Db	177046	GGCCTTCTTCCCTCACTAAGAGCCCTTATTTGAATTCACTGTGTGTGAGCCCTAGCCCTC	176987
OY	5285	CATTCTGCACATTCGCCACTCCAGCCCTTCCAAAGCAGAGACTAGGTGCCCTGCATTC	5344
Db	176986	CATTCTGCACATTCGCCACTCCAGCCCTTCCAAAGCAGAGACTAGGTGCCCTGCATTC	176927
OY	5345	CACCCAAAGGTGGATTTGGCTCTTCTTAAGGCGGTGACTGTGCACATCACGCAATACT	5404
Db	176926	CACCCAAAGGTGGATTTGGCTCTTCTTAAGGCGGTGACTGTGCACATCACGCAATACT	176867
OY	5405	GTTGCTCTCAAGACACACGCTGCCATTTTCTTCAACTGAGGGCTCAAAACTCTGGA	5464
Db	176866	GTTGCTCTCAAGACACACGCTGCCATTTTCTTCAACTGAGGGCTCAAAACTCTGGA	176807
OY	5465	CAAGTTGCTGGCTCTGTAGACACATTTTCTGTGAGCGTGGCCACTGAAAGGGGCCAG	5524
Db	176806	CAAGTTGCTGGCTCTGTAGACACATTTTCTGTGAGCGTGGCCACTGAAAGGGGCCAG	176747
OY	5525	CCTGAGGAACCTGGCTTTTCTTTAAAGCCAGGCCCACTTACATTAACACTTTCAG	5584
Db	176746	CCTGAGGAACCTGGCTTTTCTTTAAAGCCAGGCCCACTTACATTAACACTTTCAG	176687
OY	5585	GGTCACTGGAAACAGTGAAGTGCATTTGTGAAGCCTACTGCATGCCACCACTGCTC	5644
Db	176686	GGTCACTGGAAACAGTGAAGTGCATTTGTGAAGCCTACTGCATGCCACCACTGCTC	176627
OY	5645	ATCCACGTGTCTGCATGCTTACGAGGAAGGCCAGGCATGCGAGACTGCTCTAATG	5704
Db	176626	ATCCACGTGTGTCTGCATGCTTACGAGGAAGGCCAGGCATGCGAGACTGCTCTAATG	176567
OY	5705	CTGTGTCTATTGCAACAGAGGAAAGGCTTCAAGGAAGTCAACTGGGACAAGCACAG	5764
Db	176566	CTGTGTCTATTGCAACAGAGGAAAGGCTTCAAGGAAGTCAACTGGGACAAGCACAG	176507
OY	5765	CCCAACGGACATATGCTTGTGTAAGGTATTAGACACTGTGTGTGTGATCTGCAGTGTCTT	5824
Db	176506	CCCAACGGACATATGCTTGTGTAAGGTATTAGACACTGTGTGTGTGATCTGCAGTGTCTT	176447

QY	5825	CACGCGAAATATATTTATTCATGTGCAGATACCTTTTAGGTGGCATTTATTCATTTCCGT	5884
Db	176446	CACGCGAAATATTTATTCATGTGCAGATACCTTTTAGGTGGCATTTATTCATTTCCGT	176387
QY	5885	GCTTTAATTAACAATGTACCACAAAAACAAGTATCAAGCTGTTTAAGTGTCCGGTAC	5944
Db	176386	GCTTTAATTAACAATGTACCACAAAAACAAGTATCAAGCTGTTTAAGTGTCCGGTAC	176327
QY	5945	TTGTCCCTCGGTTCAGTAGAGGCCCGGGTTTCCAGTTGTGTGACTGTACAGAGCTACAGA	6004
Db	176326	TTGTCCCTCGGTTCAGTAGAGGCCCGGGTTTCCAGTTGTGTGACTGTACAGAGCTACAGA	176267
QY	6005	TGGGCTCAGCAGATGTCTCTTTAATTTTGTGATGATACAGAAAGCCAGGCTTTGGGATAC	6064
Db	176286	TGGGCTCAGCAGATGTCTCTTTAATTTTGTGATGATACAGAAAGCCAGGCTTTGGGATAC	176207
QY	6065	AAGTCTTTCCTCTTCATTTATGATGCCGTCGACTGTGTAGACAGATGTTTTGTCCGGA	6124
Db	176206	AAGTCTTTCCTCTTCATTTATGATGCCGTCGACTGTGTAGACAGATGTTTTGTCCGGA	176147
QY	6125	ATTAATAATATATGCTTGGAGTCTGCACA	6153
Db	176146	ATTAATAATATATGCTTGGAGTCTGCACA	176118

RESULT 7				
LOCUS	AKO26622			
DEFINITION	Homo sapiens CDNA: FLJ22969 fls, clone KAT10759.	3324 bp	mRNA	linear
ACCESSION	AKO26622			PRI 29-SEP-2000
VERSION	AKO26622.1	GI:10439514		
KEYWORDS	Oligo capping; fls (full insert sequence).			
SOURCE	Homo sapiens signet-ring cell carcinoma cell_line:KATO III CDNA to mRNA, clone_1lb:KAT clone:KAT10759.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi; Mammalia: Eutheria; Primates; Catarrhini; Homnidae; Homo.			
AUTHORS	1 (sites)			
TITLE	Watanabe,K., Kumagai,A., Itakura,S., Yamazaki,M., Tashiro,H., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano.S.			
JOURNAL	NEDO human CDNA sequencing project			
REFERENCE	Unpublished			
AUTHORS	2 (bases 1 to 3224)			
TITLE	Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.			
JOURNAL	Direct Submission			
COMMENT	Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)			
	NEDO human CDNA sequencing project supported by Ministry of International Trade and Industry of Japan: cDNA full insert sequencing. Research Association for Biotechnology: cDNA library construction. S - & 3'-end one pass sequencing: Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).			

URES	source	Location/Qualifiers
1..3224		/organism="Homo sapiens" /db_xref="taxon:9606" /clone="KAT10739" /cell_line="KATO III" /cell_type="signet-ring cell carcinoma" /clone_1ib="KAT" /note="cloning vector pME18FL3" 111..2621
CDS		/note="unnamed protein product" /codon_start=1 /protein_id="BAB1511.1" /db_xref="GI:10439515"

Oy	2033	AGGAGGCGCTGCCTACATCTTTCCTTTAAAGAGGGAGCGCGCTGGTCGTCGACAGACAGGGCGC	20922
Db	1861	AGGTGGCGCTGCCTACATCTTTCCTTTAAAGAGCGAGCGCGCTGGTCGTCGACAGACAGGGCGC	19202
Oy	2093	CATTCAATGATCATCCAGAGACAGCGGACCCGGGCTAGAGAGATCTTTCAGCTGGACGAG	21522
Db	1921	CATTCAATGATCATCCAGAGACAGCGGACCCGGGCTAGAGAGATCTTTCAGCTGGACGAG	19802
Oy	2153	ATGTGCTCCCCAGCCAAAGTTCCACCATTCACAGCTTTCGGGTCAACATCTCTAACTGCA	22122
Db	1981	ATGTGCTCCCCAGCCAAAGTTCCACCATTCACAGCTTTCGGGTCAACATCTCTAACTGCA	20402
Oy	2213	GGCCCCAGAGGGGGAAACACACTGAGACTGCTGCTCTTCGGGTGACACTTACCTTCCCAAGAGATG	22722
Db	2041	GGCCCCAGAGGGGGAAACACACTGAGACTGCTGCTCTTCGGGTGACACTTACCTTCCCAAGAGATG	21002
Oy	2273	TGGACTTGCATCTCATCTCATTCGACAGCGGTGGGAGGTGAGATCTTACTGCTGTGGCC	23322
Db	2101	TGGACTTGCATCTCATCTCATTCGACAGCGGTGGGAGGTGAGATCTTACTGCTGTGGCC	21362
Oy	2333	TGCGGCTCATCATTTTCTGTGTGAAAAAGAAAAAGAAACAACAAAGGGCCCCGCTG	23922
Db	2161	TGCGGCTCATCATTTTCTGTGTGAAAAAGAAAAAGAAACAACAAAGGGCCCCGCTG	22202
Oy	2393	TGGGATCTTCAAAAGGCAATCAATACAGATGTCGGCAGGAGCAAAAGTTTCAGA	24322
Db	2221	TGGGATCTTCAAAAGGCAATCAATACAGATGTCGGCAGGAGCAAAAGTTTCAGA	22802
Oy	2453	AAGGGCGAAAGGACAATGACTCCCATGTGATTCATCGAGACACACATGATATG	25122
Db	2281	AAGGGCGAAAGGACAATGACTCCCATGTGATTCATCGAGACACACATGATATG	23402
Oy	2513	GGCATCTGCTACAGAGATTCACAGCGGCTCCTTCTGCGACGACAGAGTGGACATACGGC	25722
Db	2341	GGCATCTGCTACAGAGATTCACAGCGGCTCCTTCTGCGACGACAGAGTGGACATACGGC	24002
Oy	2573	CGTTCACAGGGACCATGSGGGGTGTGTCTCCGCCCCACCAACAAATGTCTCAAGGGCC	26322
Db	2401	CGTTCACAGGGACCATGSGGGGTGTGTCTCCGCCCCACCAACAAATGTCTCAAGGGCC	24602
Oy	2633	CAACTGCAAAAGTTGGCCACGAGAGGACCACTTCCTGCTCCCTCATAGTCTGAGATG	26922
Db	2461	CAACTGCAAAAGTTGGCCACGAGAGGACCACTTCCTGCTCCCTCATAGTCTGAGATG	25202
Oy	2693	AACCGTACACCTTCTCCCATCCCAACACATGSGGATGTAAAGCAGCAAGACACACATTC	27522
Db	2521	AACCGTACACCTTCTCCCATCCCAACACATGSGGATGTAAAGCAGCAAGACACACATTC	25802
Oy	2753	CGTTACGACACACTGACAGAGCCCATGAGCGACACAAATATCTGATCATTTCCAGACGC	28122
Db	2581	CGTTACGACACACTGACAGAGCCCATGAGCGACACAAATATCTGATCATTTCCAGACGC	26402
Oy	2813	TTTCTCTAGATTTCATTAAGAGGGGACCTGACACCCGTCCTGTTCCTAACACAAATTC	28722
Db	2641	TTTCTCTAGATTTCATTAAGAGGGGACCTGACACCCGTCCTGTTCCTAACACAAATTC	27002
Oy	2873	CTAAAGAAGAGGAATTTCACAGAGGACACGAGAGTTTTCTGTGACACCGCCCACTT	29322
Db	2701	CTAAAGAAGAGGAATTTCACAGAGGACACGAGAGTTTTCTGTGACACCGCCCACTT	27602
Oy	2933	CACATCTGCTCAGTGGACTCATTTCTAAAGGCGACACATTTAAAAATGATGAATTCACATTCG	29922
Db	2761	CACATCTGCTCAGTGGACTCATTTCTAAAGGCGACACATTTAAAAATGATGAATTCACATTCG	28202
Oy	2993	GATACAGTCAATGACAGTCAATGTGTCCTCAACTTATGAGTGTGGGTATAGCAGCGCTGA	30522
Db	2821	GATACAGTCAATGACAGTCAATGTGTCCTCAACTTATGAGTGTGGGTATAGCAGCGCTGA	28802
Oy	3053	ATGAGAGAGAGAGCGCTTGATCACTAGCATATGAGTTTCACACAAAGCCCTGGATTCCAGAG	31122
Db	2881	ATGAGAGAGAGAGCGCTTGATCACTAGCATATGAGTTTCACACAAAGCCCTGGATTCCAGAG	29402
Oy	3113	TGTTTAAACAGAGCTGCCTCTTTCACAGACACATGTTCCATTTCCAAAGAGACTCACTGA	31722

Db	2941	TGTTAAACAGAGCGCTGGCCCTCTTTCAGACAACAGTTCCATTCCAAAGAGAGCTACGTGA	3000
Qy	3173	GTCCTCTACTCTCACTGGGGGCCACAGATATAAAACGACAAATGCCCCCTTTATATATAT	3232
Db	3001	GGTCCCTACTCTCACTGGGGGCCACAGATATAAAACGACAAATGCCCCCTTTATATATAT	3060
Qy	3233	TTATTTGGTGTGCTCTGTGTTATTTAAGAGATCAAAATGTATAACCACTAGCTCTTTTCAC	3292
Db	3061	TTATTTGGTGTGCTCTGTGTTATTTAAGAGATCAAAATGTATAACCACTAGCTCTTTTCAC	3120
Qy	3293	CTGACTTTAGTATACTATCTACTAAGCTGGTTGGATGGCTGGGTGTACTTCTCTCTAC	3352
Db	3121	CTGACTTTAGTATACTATCTACTAAGCTGGTTGGATGGCTGGGTGTACTTCTCTCTAC	3180
Qy	3353	CGCTAGATAAACGCTGGCTGTCCCCCA	3380
Db	3181	CGCTAGATAAACGCTGGCTGTCCCCCA	3208

RESULT 8	AK026028	LOCUS	DEFINITION	VERSION	KEYWORDS	SOURCE
	AK026028	2113 bp	linear			
		AK026028	Homo sapiens CDNA: FLJ22375 fis, clone HRC6785.			
		AK026028	AK026028.1 GI:10438737			
			Oligo capping: fis (full insert sequence).			
			Homo sapiens primary human renal epithelial cells CDNA to mRNA,			

ORGANISM

REFERENCE
AUTHORS
1 (sites)
Kawabata, A., Hikiji, T., Kobatake, N., Inagaki, H., Ikema, Y., . . .

TITLE	NEDO human cDNA sequencing project
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 2113)
AUTHORS	Sugano, S., Suzuki, Y., Ota, T., Ohyashiki, M., Nishii, T., Isega, T., Shihabara, T., Tanaka, T. and Nakamura, Y.

JOURNAL
Submitted (23-AUG-2000) Sumio Sugano, Institute of Medical Science,
University of Tokyo, Laboratory of Genome Structure Analysis, Human
Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,
Japan (E-mail: cdna@elms.u-tokyo.ac.jp, Tel: 81-3-5449-5286,
Fax: 81-3-5449-5416)
NEO human cDNA sequencing project supported by Ministry of

sequencing: Research Association for Biotechnology; cDNA library construction, 5' - & 3' - end one pass sequencing: Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).

source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HRC06785"
/cell_type="primary human renal epithelial cells"
/clone_id="HRC"
/note="cloning vector PMEL18Fv3"

```

Query Match	30.4%	Score 1875.8	DB 9	Length 2113
Best Local Similarity	99.8%	Pred. NO. 0		
Matches 1888	Conservative	0	Mismatches 2	Indels 1
				Gaps 1

Qy	2448	TAGAAAGGGCGAAAGACAATGACTCCCATGTGTATGCAGTATCATCAGGACACCATGGT	2507
Db	1	TAGAAAGGGCGAAAGACAATGACTCCCATGTGTATGCAGTATCATCAGGACACCATGGT	60

Qy	2508	ATATGGGCAATCTGCTACAGGAATTCACGCGGCTCTTCTCTGACGCCAGAGGTGGACACTA	2567
Dp	61	ATATGGGCAATCTGCTACAGGAATTCACGCGGCTCTTCTCTGACGCCAGAGGTGGACACTA	120
Qy	2568	CCGGCCGTTCCAGGGGACCAATGGGGGTCTGTCTCTCTCCACCCACCAATATGTCTCAG	2627
Dp	121	CCGGCCGTTCCAGGGGACCAATGGGGGTCTGTCTCTCTCCACCCACCAATATGTCTCAG	180
Qy	2628	GGCCCCAACTGCAAAAGTTGGACCACTAGAGAGCCACCTCTCGCTCCCTCCGTAAGTCTGA	2687
Dp	181	GGCCCCAACTGCAAAAGTTGGACCACTAGAGAGCCACCTCTCGCTCCCTCCGTAAGTCTGA	240
Qy	2688	GAGTGAACCGTACACCTTCTCCCATCCCAAAATGGGGATGTAGACAGCAGACACAGA	2747
Dp	241	GAGTGAACCGTACACCTTCTCCCATCCCAAAATGGGGATGTAGACAGCAGACACAGA	300
Qy	2748	CATTCCCTTACTGAAACCTAGAGAGCCAAATGGAGCCAGAGAAATTAATGTATCCATTCCA	2807
Dp	301	CATTCCCTTACTGAAACCTAGAGAGCCAAATGGAGCCAGAGAAATTAATGTATCCATTCCA	360
Qy	2808	GAGCCTTTCGTGAGTTTCATAAAGACAGAGGCACAGACACCCTCGCTCCCTCCGTAAGCAG	2867
Dp	361	GAGCCTTTCGTGAGTTTCATAAAGACAGAGGCACAGACACCCTCGCTCCCTCCGTAAGCAG	420
Qy	2868	AAATCTCTAAAGAAAGGAATTTATACAGAAAGAAACAGCAGAGGTTTCTGTGACACCGCC	2927
Dp	421	AAATCTCTAAAGAAAGGAATTTATACAGAAAGAAACAGCAGAGGTTTCTGTGACACCGCC	480
Qy	2928	AACTTTCACATTTGCTCANTGACATCTTTAAGGGCAAGACATTTGAAAAATGATGAATTCGA	2987
Dp	481	AACTTTCACATTTGCTCANTGACATCTTTAAGGGCAAGACATTTGAAAAATGATGAATTCGA	540
Qy	2988	ATTCGATACAGTCATACAGCTCATGTGCTCTCAACTAGGCTGTGGGGTTTACGACAGC	3047
Dp	541	ATTCGATACAGTCATACAGCTCATGTGCTCTCAACTAGGCTGTGGGGTTTACGACAGC	600
Qy	3048	CTGTAAATGAAGAGAGAGAGCGGCTGAGTCAACCTAGATAGGTTTGCAGCAAGCCCTGAGTT	3107
Dp	601	CTGTAAATGAAGAGAGAGAGCGGCTGAGTCAACCTAGATAGGTTTGCAGCAAGCCCTGAGTT	660
Qy	3108	CAGAGTGTTTAAACAGAGGCTTGGCCCTTTCAGAGACAAGTTCCAAATTCGAAGAGCCTA	3167
Dp	661	CAGAGTGTTTAAACAGAGGCTTGGCCCTTTCAGAGACAAGTTCCAAATTCGAAGAGCCTA	720
Qy	3168	CCTAGAGTCCCTACTCTCAGCTGGGGTCCCGAGGATTAAGAACACATGTGCTTTTATTT	3227
Dp	721	CCTAGAGTCCCTACTCTCAGCTGGGGTCCCGAGGATTAAGAACACATGTGCTTTTATTT	780
Qy	3228	ATTATTTATTTGGTGGTCTGTGTATTTAAGAGATCAAAATGTATTAACCACTACTACTT	3287
Dp	781	ATTATTTATTTGGTGGTCTGTGTATTTAAGAGATCAAAATGTATTAACCACTACTACTT	840
Qy	3288	TTTCACTGACTTACTAATTAACATATACATCTGATTTGGATGCTGGGTTGTACTCTTA	3347
Dp	841	TTTCACTGACTTACTAATTAACATCTGATTTGGATGCTGGGTTGTACTCTTA	900
Qy	3348	CTGACGCGCTAGATTAAGGTGTGCTGTCCGCCACGAGTGGGGGAATTAATTTCAAACTCTC	3407
Dp	901	CTGACGCGCTAGATTAAGGTGTGCTGTCCGCCACGAGTGGGGGAATTAATTTCAAACTCTC	960
Qy	3408	CAACACAGAAAAAGATGTGTGTATTAGACAGCAATTACACATATCTGTTTGATTAAGAGA	3467
Dp	961	CAACACAGAAAAAGATGTGTGTATTAGACAGCAATTACACATATCTGTTTGATTAAGAGA	1020
Qy	3468	CTTCTCTATTTCTCTAGGTGCGTTCTGTGTATCCATTGTGGAATTCATCTTGATATCC	3527
Dp	1021	CTTCTCTATTTCTCTAGGTGCGTTCTGTGTATCCATTGTGGAATTCATCTTGATATCC	1080
Qy	3528	ATTGTCTATATAGTCATAGCAATTAAGAAATTTCCCAAGTTTCCATGTGGGGTCTCTCT	3587
Dp	1081	ATTGTCTATATAGTCATAGCAATTAAGAAATTTCCCAAGTTTCCATGTGGGGTCTCTCT	1140
Qy	3588	AGCTGCAGCATACTTTGACATTTTAAAGAAATTTAGAAATTAATTCATCTCTTAATA	3647

DB	1141	ACGTCACACAAATACCTTTAGACATTTAAAGAAATTTAAGAAATATCTTCATCTCTTAAAA	1200
QY	3648	ATGTTTAATATATATACCAAAAGTGGCCCCGCGATGATGTTTGTGTTGGCCACAGCAAC	3707
Db	1201	ATGTTTAAATATATATACCAAAAGTGGCCCCGCGATGATGTTTGTGTTGGCCACAGCAAC	1260
QY	3708	CATTACTTGGTAGCTTAAAAAACMACACATATAGCTTATAGTCCCTGGGGATCAGAAATCCAA	3767
Db	1261	CATTACTTGGTAGCTTAAAAAACMACACATATAGCTTATAGTCCCTGGGGATCAGAAATCCAA	1320
QY	3768	AATGATATGCTCTGTAATGAAATCAAGGTGTGACAGAGCTGTGCTCTTCTGAAAGCTC	3827
Db	1321	AATGATATGCTCTGTAATGAAATCAAGGTGTGACAGAGCTGTGCTCTTCTGAAAGCTC	1380
QY	3828	TAGGAGAAAGCCGGGTCCCTGGCATTTCAAGCTTTAGAGCTGGCTCATTTCCAGAGCT	3887
Db	1381	TAGGAGAAAGCCGGGTCCCTGGCATTTCAAGCTTTAGAGCTGGCTCATTTCCAGAGCT	1440
QY	3888	CCAGTGGCTGTGCAAGCTTTTCTCATGATGATCAGTGTGACACTGGCCCTCCACTTCC	3947
Db	1441	CCAGTGGCTGTGCAAGCTTTTCTCATGATGATCAGTGTGACACTGGCCCTCCACTTCC	1500
QY	3948	CTCTTTGACTTACAAAGCCACACAGGAAGATCAGAGTAATCTCTCCATCTTAAAGATCTT	4007
Db	1501	CTCTTTGACTTACAAAGCCACACAGGAAGATCAGAGTAATCTCTCCATCTTAAAGATCTT	1560
QY	4008	TCATATCTCTGGAAGAGCTTTGGCATGCAAGAACAAATATAGCCACAGTGGGGATTTAGG	4067
Db	1561	TCATATCTCTGGAAGAGCTTTGGCATGCAAGAACAAATATAGCCACAGTGGGGATTTAGG	1620
QY	4068	ACCAGAGACATCTTTGGGTGCTGTATTTCTGCTTACACACACTTCTCCACTGACTGCC	4127
Db	1621	ACCAGAGACATCTTTGGGTGCTGTATTTCTGCTTACACACACTTCTCCACTGACTGCC	1680
QY	4128	ACAGAGAGGCTTACAAAATGATCTGGCCGACAGGAGTGTGTTTAACTTGGGACTCT	4187
Db	1681	ACAGAGAGGCTTACAAAATGATCTGGCCGACAGGAGTGTGTTTAACTTGGGACTCT	1740
QY	4188	AACACTT-AAAAAAACCCCAAGATCAGAAAGATCTGGCCATGCTGGGGGTCACATTTCTACC	4246
Db	1741	AACACTTAAAAAAACCCCAAGATCAGAAAGATCTGGCCATGCTGGGGGTCACATTTCTACC	1800
QY	4247	TAGCAACAACCTGGCTGAGAGCTGGGACACACAGCTCTGCTTTAGAAGGGGTGTCCACTTACC	4306
Db	1801	TAGCAACAACCTGGCTGAGAGCTGGGACACACAGCTCTGCTTTAGAAGGGGTGTCCACTTACC	1860
QY	4307	CAGTCCACACAGCCACACACTAGGCCATATC 4337	
Db	1861	CAGTCCACACAGCCACACACTAGGCCATATC 1891	
RESULT 9			
LOCUS	AK026187	1241 bp	mRNA linear PRI 29-SEP-2000
DEFINITION	Homo sapiens cDNA: FLJ22534 fls, clone HRC13020.		
ACCESSION	AK026187		
VERSION	AK026187.1	GI:10438956	
KEYWORDS	oligo capping; fls (full insert sequence).		
SOURCE	Homo sapiens primary human renal epithelial cells cDNA to mRNA, clone_11b:HRC clone:HRC13020.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (siles)		
	Kawabata,A., Hiki,J., Kobatake,N., Inagaki,H., Ikema,Y., Okamoto,S., Ohtani,I., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shiba-hara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.		
TITLE	NEO human cDNA sequencing project		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 1241)		
AUTHORS	Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shiba-hara,T., Tanaka,T. and Nakamura,Y.		

FEATURES	source	Location/Qualifiers
COMMENT		<p>Direct Submission</p> <p>Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail: cdhale@ims.u-tokyo.ac.jp, Tel: 81-3-5449-5286, Fax: 81-3-5449-5416)</p> <p>NEO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan: cDNA full insert sequencing: Research Association for Biotechnology: cDNA library construction, 5'- 8' end one pass sequencing: Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).</p>
FEATURES	source	<p>Location/Qualifiers</p> <p>1. .1241</p> <p>/organism="Homo sapiens"</p> <p>/db_xref="taxon:9606"</p> <p>/clone="HRC13020"</p> <p>/cell_type="primary human renal epithelial cells"</p> <p>/clone_11p="HRC"</p> <p>/note="cloning vector pWE18FLJ3"</p> <p>65. .>1241</p> <p>/note="unnamed protein product"</p> <p>/codon_start=1</p> <p>/protein_id="BAB15388.1"</p> <p>/db_xref="GI:10438957"</p> <p>/translation="MAGNFNLISLQGGQDDAOSPILRLQFVLVQHPQNESNKIYVVDLSNRASLITETPRPVKSRKREYVGCFCVLESFSSNLITLSSKRIISPLCDLRLRMNVETISCTDHRVCOAKSYSLQVADLIHLVEHLHPFWKILVPRDLSTLVPARORLQVETKSTDRKQVARSATPSODIXRSGCPGSGIKQIQKQINVSITLTPARSFQDEASRQGLITVSFIPIFKKEGVFTYTPDKSKVYVLRTPMDGLSLVSNNISVPRDDVLTFFPKERSGVVCCQGRFAMVLIQDRTAEEVLSLDEVLLKPSFHHHSFWVFNISCSPTSGKQDLDFPSVLTLPRTVDLYLLAAVGGVLLLSALGLICQVKKKKK"</p>
CDS		
BASE COUNT	296 a	380 c 313 g 252 t
ORIGIN		
Query Match	20.0%	Score 1232.6; DB 9; Length 1241;
Best Local Similarity	99.7%	Pos. No. 8.8e-249;
Matches 1235; Conservative	4;	Mismatches 4; Indels 0; Gaps 0;
OY	1133	ACATCCGGGGCTCCACCAACCCCGAGGAGTTCAAGCTGGAGGACAAGCAGCTGGGA 1192
Db	3	ACATCCGGGGCTCCACCAACCCCGAGGAGTTCAAGCTGGAGGACAAGCAGCTGGGA 62
OY	1193	ACATGGCGGGGAATTCACCTCTCTGTGCAGAGCGTGTGACCAAGATGCCAAATGCCAG 1252
Db	63	ACATGGCGGGGAATTCACCTCTCTGTGCAGAGCGTGTGACCAAGATGCCAAATGCCAG 122
OY	1253	GGATCCTCCGGGTGAGTTCAGAGTTTGGTCCAAATCCCAAAATATAAGCAATTAAA 1312
Db	123	GGATCCTCCGGGTGAGTTCAGAGTTTGGTCCAAATCCCAAAATATAAGCAATTAAA 182
OY	1313	TCTAGTGGTTGACTTGAATGATGAGCAGGCATGTCACTACATCGAGGCACGGCCG 1372
Db	183	TCTAGTGGTTGACTTGAATGATGAGCAGGCATGTCACTACATCGAGGCACGGCCG 242
OY	1373	TCAACACAGACCCGCAGATTGTCTCCCTGGCTGTTTCGTGTGATAGATCTGGACCTGCA 1432
Db	243	TCAACACAGACCCGCAGATTGTCTCCCTGGCTGTTTCGTGTGATAGATCTGGACCTGCA 302
OY	1433	GTAGAACCTCAACCTGACATCTGGCTCCAAACAAATAATCTCTTGTGATGATC 1492
Db	303	GTAGAACCTCAACCTGACATCTGGCTCCAAACAAATAATCTCTTGTGATGATC 362
OY	1493	TGACACGCTGTGTGATGATGTGAAAAAACCATTAAGCTGACAGACACCGGTACTGCC 1552
Db	363	TGACACGCTGTGTGATGATGTGAAAAAACCATTAAGCTGACAGACACCGGTACTGCC 422
OY	1553	AAAGAAATCTCTACTACTCCAGGTGCCAAGTGACATCTCCACCTGGCTGTGAGCTGC 1612
Db	423	AAAGAAATCTCTACTACTCCAGGTGCCAAGTGACATCTCCACCTGGCTGTGAGCTGC 482

OY	1613	ATGACTTTCCTCGGAAGCTGCTGGTGGCCCAAGAGACAGGCTACAGCTGGTGGTGGCCAG	1672
Db	483	ATGACTTTCCTCGGAAGCTGCTGGTGGCCCAAGAGACAGGCTACAGCTGGTGGTGGCCAG	542
OY	1673	CCGAGAACCTCGACAGAGCATACACAGAAAGCCCTGCAACACAGCTTCAGTACCTCG	1732
Db	543	CCGAGAGGCTCGAGAGAGCATACACAGAAAGCCCTGCAACACAGCTTCAGTACCTCG	602
OY	1733	TGGCCATGTCATACACAGGACAGGACCTGTACTTGGCTCCTTCTGCCCCGAGGCTCTA	1792
Db	603	TGGCCATGTCATACACAGGACAGGACCTGTACTTGGCTCCTTCTGCCCCGAGGCTCTA	662
OY	1793	TCGAACATATCAGGTGTAAGAGAGAATCATCTGGGTGACCTTCGACACCTTTCGCCACACT	1852
Db	663	TCGAACATATCAGGTGTAAGAGAGAATCATCTGGGTGACCTTCGACACCTTTCGCCACACT	722
OY	1853	TCGAACAGAGAGGCTCCACGAGAGGCTGTAGACGCTGTCCCTTTATACCTATTTTCAAGAG	1912
Db	723	TCGAACAGAGAGGCTCCACGAGAGGCTGTAGACGCTGTCCCTTTATACCTATTTTCAAGAG	782
OY	1913	AAGGCGTTTTCACGCTGACCCCTGACACAAAAAGCAGAGTCTACCTGAGGACCCCAACT	1972
Db	783	AAGGCGTTTTCACGCTGACCCCTGACACAAAAAGCAGAGTCTACCTGAGGACCCCAACT	842
OY	1973	GGGAGCCGGGGGCTGCATACCCCTGACACCTGTGTCTGGACATACAGGCTGGCCAGAGAC	2032
Db	843	GGGAGCCGGGGGCTGCATACCCCTGACACCTGTGTCTGGACATACAGGCTGGCCAGAGAC	902
OY	2033	AGGTGGCTTGGCTGACTTTCTTTAAGAGACGAGCGGCGTGGCTGGCCACAGAGGCGCG	2092
Db	903	AGGTGGCTTGGCTGACTTTCTTTAAGAGAGGAGGCGGCTGGCTGGCCACAGAGGCGCG	962
OY	2093	CATTATATATCATCCAGAGACAGCGGACCCGGGCTGAGAGATCTTCAGCTTGCACAGG	2152
Db	963	CATTATATATCATCCAGAGAGAGCGGACCCGGGCTGAGAGATCTTCAGCTTGCACAGG	1022
OY	2153	ATGTCTCTCCCAAGGCAAGCTTCCACATCATCAGCTTTGGGTGAACATCTCTAAGTCA	2212
Db	1023	ATGTCTCTCCCAAGGCAAGCTTCCACATCATCAGCTTTGGGTGAACATCTCTAAGTCA	1082
OY	2213	GCCCCACAGAGCGGCAAGCAGCTAGACCTGCTCTCTCGGTGACACTTACCCCAAGGACTG	2272
Db	1083	GCCCCACAGAGCGGCAAGCAGCTAGACCTGCTCTCTCGGTGACACTTACCCCAAGGACTG	1142
OY	2273	TGGACTTACGTGTCACTCTCATCGCAGCGGTGGAGGTGAGTCTTACTGCTGTGCCC	2332
Db	1143	TGGACTTACGTGTCACTCTCATCGCAGCGGTGGAGGTGAGTCTTACTGCTGTGCCC	1202
OY	2333	TCGGGCTCATCATTTGCTGTGTAAGAAAAAGAAAAAGA 2371	
Db	1203	TCGGGCTCATCATTTGCTGTGTGAAAAAAGAAAAA 1241	
RESULT 10			
LOCUS	BC021099	1410 bp	mRNA
DEFINITION	Homo sapiens, similar to hypothetical protein FLJ22969, clone		
ACCESSION	MGC:31813 IMAGE:4590554, mRNA, complete cds.		
VERSION	BC021099		
KEYWORDS	MGC.		
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
TITLE	1 (bases 1 to 1410)		
JOURNAL	Strausberg,R. Direct Submision		
REMARK	Submitted (03-JAN-2002) National Institutes of Health, Mammalian		
	Gene Collection (MGC), Cancer Genomics Office, National Cancer		
	Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,		
	USA		
	NIH-MGC Project URL: http://mgc.ncl.nih.gov		

COMMENT

Contact: MGC help desk
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: DCTD/DRP
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Genome Sequence Centre,
 BC Cancer Agency, Vancouver, BC, Canada
 info@gscc.bc.ca
 Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
 Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
 Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
 Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
 Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline
 Schein, Duane Smellus, Michael Smith, Lorraine Spence, Jeff Stott,
 Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
 George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Series: IRML Plate: 40 Row: f Column: 14
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 12383093.
 Location/Qualifiers

FEATURES

source

1. .1410
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="MGC:31813 IMAGE:4590554"
 /tissue_type="Kidney, renal cell adenocarcinoma"
 /clone_1lb="NH_MGC_14"
 /lab_host="DH10B-R"
 /note="Vector: pOT87"
 136. .1167
 /codon_start=-1
 /product="Similar to hypothetical protein FLJ22969"
 /protein_id="AAH21099.1"
 /db_xref="gi:18088979"
 /translation="MAGLNGVSIALLGLVLLGAARLPGRGAFETALPRENSITVLI
 KIGTPILAKPCYIVISKRHTMISKRSERYFTFSCSPENHRYETLOKINIDCSG
 PCPEEVOLQPTSLPTLNRTIMVNAHKSIGLHLOPSIFRLQIGESCPDGV
 HSIQRIDATVVRIGTFCNSGTVSRIMQEGVMALHLPFRPVRNYSFINRSRISIK
 RCIIESEVEGEGSATLMSANYPEGPEDELTWQFVVAHRLHRSVSEFINLSNCR
 KEEREVYIPGSTNPVEVFKLBDKQGNMAGNPNLSLQCDQDAGSPGILRLQFQVLV
 QHPONSSS"

CDS

BASE COUNT 321 a 395 c 370 g 324 t
 ORIGIN

Query Match 18.8%; Score 1159.8; DB 9; Length 1410;
 Best Local Similarity 99.8%; Pred. No. 1.8e-233;
 Matches 1161; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 148 GGGCGGGGCTCGGGCGGTCGCCCGCGGCGCAGGTGAGTGAAGCGGCGGAGCGCAGCT 207
 |||||
 DB 1 GGGCGGGGCTCGGGCGGTCGCCCGCGGCGCAGGTGAGTGAAGCGGCGGAGCGCAGCT 60
 QY 208 GGGCGGGGCTCGGGCGGTCGCCCGCGGCGCAGGTGAGTGAAGCGGCGGAGCGCAGCT 267
 |||||
 DB 61 GGGCGGGGCTCGGGCGGTCGCCCGCGGCGCAGGTGAGTGAAGCGGCGGAGCGCAGCT 120
 QY 268 AGCGCTCCGCGAGTCAATGCGCGGCTGAATCGGGGTCCTATCCGACATGCTAGGGGTT 327
 |||||
 DB 121 AGCGCTCCGCGAGTCAATGCGCGGCTGAATCGGGGTCCTATCCGACATGCTAGGGGTT 180
 QY 328 CTGCTCTGGGTCGGCGGCGCTGCCCGCGGCGGCGCAGAGCTTTTGAGATGCTCTGCCA 387
 |||||
 DB 181 CTGCTCTGGGTCGGCGGCGCTGCCCGCGGCGGCGCAGAGCTTTTGAGATGCTCTGCCA 240
 QY 388 CGAGAAAGCAATTCAGTTCTCATAAAGCTGGGACCGCGGCTGCTGGGAAAAACC 447
 |||||
 DB 241 CGAGAAAGCAATTCAGTTCTCATAAAGCTGGGACCGCGGCTGCTGGGAAAAACC 300
 QY 448 TGTATACATGCTATTTCTAAAGACATATTAACCATGTTGTCATCAAGTCTGGAGAAGA 507
 |||||
 DB 301 TGTATACATGCTATTTCTAAAGACATATTAACCATGTTGTCATCAAGTCTGGAGAAGA 360

QY 508 ATAGCTTTACCTTTAGCTCCGAGAGTCGAGATCACTTTGTCATAGAGATCCAGAA 567
 |||||
 DB 361 ATAGCTTTACCTTTAGCTCCGAGAGTCGAGATCACTTTGTCATAGAGATCCAGAA 420
 QY 568 AATATTGACTGTATGTCAGAGCCCATGCTCTTTGGGAGGTTGACGTTCAAGCCCTGACA 627
 |||||
 DB 421 AATATTGACTGTATGTCAGAGCCCATGCTCTTTGGGAGGTTGACGTTCAAGCCCTGACA 480
 QY 628 TCGTTGTGGCTACCTCAACAGAACTTTTCATCTGGAGTCAAAAGTCAATAGACATC 687
 |||||
 DB 481 TCGTTGTGGCTACCTCAACAGAACTTTTCATCTGGAGTCAAAAGTCAATAGACATC 540
 QY 688 GGTTAGAGTCGAGTTTTCATCCCTCGGCTGAGAGAGATCCGCTCGGGGTGAGAGCTGC 747
 |||||
 DB 541 GGTTAGAGTCGAGTTTTCATCCCTCGGCTGAGAGAGATCCGCTCGGGGTGAGAGCTGC 600
 QY 748 CCAAGAGGATCACTCACTCATCAGCGGCGGATGATCCACCGGTGAGAGATCGA 807
 |||||
 DB 601 CCAAGAGGATCACTCACTCATCAGCGGCGGATGATCCACCGGTGAGAGATCGA 660
 QY 808 ACCTTTCGACGATGCGCACTGTCTCCGGATCAAGATGCAAGAGAGTGAATAATGGCC 867
 |||||
 DB 661 ACCTTTCGACGATGCGCACTGTCTCCGGATCAAGATGCAAGAGAGTGAATAATGGCC 720
 QY 868 TTACACCTCCCATGTTCCACCCGAGAAATGTCGCGGCTTCACGATGCAAAACCGCTCA 927
 |||||
 DB 721 TTACACCTCCCATGTTCCACCCGAGAAATGTCGCGGCTTCACGATGCAAAACCGCTCA 780
 QY 928 TCTATAAAGCTGCTGTCATCATGAGATCTGTGTTGAGGAGTGAAGGCTCAGAACCTTG 987
 |||||
 DB 781 TCTATAAAGCTGCTGTCATCATGAGATCTGTGTTGAGGAGTGAAGGCTCAGAACCTTG 840
 QY 988 ATGTCTGCCACATACCCAGAGGCTTCCCTGAGAGATGAGCTCATGAGTGGAGTTTGTG 1047
 |||||
 DB 841 ATGTCTGCCACATACCCAGAGGCTTCCCTGAGAGATGAGCTCATGAGTGGAGTTTGTG 900
 QY 1048 GTTCTCTGACAACTCGCGGCGGCGAGCTCTCTCTCAACTTCAACCTCTCCAACTGTGAG 1107
 |||||
 DB 901 GTTCTCTGACAACTCGCGGCGGCGAGCTCTCTCTCAACTTCAACCTCTCCAACTGTGAG 960
 QY 1108 AGAAGAGAGAGGGGTTGAATCACTATCCCGGGCTCCACACCAACCCCGAGGTTTC 1167
 |||||
 DB 961 AGAAGAGAGAGGGGTTGAATCACTATCCCGGGCTCCACACCAACCCCGAGGTTTC 1020
 QY 1168 AAGCTGAGAGACAAGACGCTGGGAACATGCGGGGAATCTCAACTCTCTGCAAGGC 1227
 |||||
 DB 1021 AAGCTGAGAGACAAGACGCTGGGAACATGCGGGGAATCTCAACTCTCTGCAAGGC 1080
 QY 1228 TGTGACCAAGATGCCCAAAATGCCAGGATCTCGGCTGCAGTTCCAAAGTTTGTGTCAA 1287
 |||||
 DB 1081 TGTGACCAAGATGCCCAAAATGCCAGGATCTCGGCTGCAGTTCCAAAGTTTGTGTCAA 1140
 QY 1288 CATCCACAAATGAAGCAATTA 1310
 |||||
 DB 1141 CATCCACAAATGAAGCAATGA 1163

RESULT 11
 AX464028
 LOCUS 1376 bp DNA linear PAT 16-JUL-2002
 DEFINITION Sequence 161 from Patent WO0140466.
 AX464028
 ACCESSION AX464028
 VERSION AX464028.1 GI:21899031
 KEYWORDS
 SOURCE human.
 ORGANISM human.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE
 1 Baker, K.P., Beresini, M., DeForge, L., Desnoyers, L., Filvaroff, E.,
 Gao, W.O., Gerlitsen, M.E., Goddard, A., Godowski, P.J., Gurney, R.L.,
 Sherwood, S., Smith, V., Stewart, T.A., Tumes, D., Watanabe, C.K.,

Wood, W.L. and Zhang, Z.
 Secreted and transmembrane polypeptides and nucleic acids encoding
 same
 JOURNAL Patent: WO 0140466-A 161 07-JUN-2001;
 Genentech Inc. (US)
 FEATURES Location/Qualifiers
 source 1..1376
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 311 a 384 c 352 g 329 t
 ORIGIN

Query Match 18.1%; Score 1117.8; DB 6; Length 1376;
 Best Local Similarity 99.8%; Pred. No. 1.3e-224;
 Matches 1119; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 190 CCAGGGGAGAGCAGCTGCGCGGCTTGGGCGCTGGGGCGCGCGCTCCCAACCGTCG 249
 |||||||
 DB 1 CCAGGGGAGAGCAGCTGCGCGGCTTGGGCGCTGGGGCGCGCGCTCCCAACCGTCG 60

OY 250 TTTTCCCGACCGAGCGAGCGCTCCGGAGTCATGGCGCGCTGAACCTGGGGCTCT 309
 |||||||
 DB 61 TTTTCCCGACCGAGCGAGCGCTCCGGAGTCATGGCGCGCTGAACCTGGGGCTCT 120

OY 310 ATGCGACCTGCTAGGGGTTCTGCTGCTGGGCGCGCGCTCCCGCGCGGGGAGAGCT 369
 |||||||
 DB 121 ATGCGACCTGCTAGGGGTTCTGCTGCTGGGCGCGCGCTCCCGCGCGGGGAGAGCT 180

OY 370 TTTAGATTGCTCTGCGACGAGAAACATATACAGTTCTCATTAAGCTGGGAGCCCG 429
 |||||||
 DB 181 TTTAGATTGCTCTGCGACGAGAAACATATACAGTTCTCATTAAGCTGGGAGCCCG 240

OY 430 ACTCTGCTGGCAAAACCTGTTACATGTCATTTCTAAAGACATATAACCATGTTGTC 489
 |||||||
 DB 241 ACTCTGCTGGCAAAACCTGTTACATGTCATTTCTAAAGACATATAACCATGTTGTC 300

OY 490 ATCAAGCTGGAGAAAGATATGTTTACCTTACCTTACCTGCGAGATCCTGAGATCATT 549
 |||||||
 DB 301 ATCAAGCTGGAGAAAGATATGTTTACCTTACCTTACCTGCGAGATCCTGAGATCATT 360

OY 550 GTCATAGAGATCCAGAAATATGACTGATGATGACAGCCCATGCTTTTGGGAGGTT 609
 |||||||
 DB 361 GTCATAGAGATCCAGAAATATGACTGATGATGACAGCCCATGCTTTTGGGAGGTT 420

OY 610 CAGCTTACGCGCTGACATGTTGTTGCTACCTCAACAGAACTTTCATCTGGAGATGTC 669
 |||||||
 DB 421 CAGCTTACGCGCTGACATGTTGTTGCTACCTCAACAGAACTTTCATCTGGAGATGTC 480

OY 670 AAAGCTATAAGAGATCGGTTTACGATGATGATGATGATGATGATGATGATGATGAT 729
 |||||||
 DB 481 AAAGCTATAAGAGATCGGTTTACGATGATGATGATGATGATGATGATGATGATGAT 540

OY 730 GGTCCGGGTAGAGATCGCGAGAGAGATCACTCACTCACTCACTCACTCACTCACTCACT 789
 |||||||
 DB 541 GGTCCGGGTAGAGATCGCGAGAGAGATCACTCACTCACTCACTCACTCACTCACTCACT 600

OY 790 ACCGTGTCAGAGATCGAGACCTTCTGACAGCAATGGCACTGTGCTCCGATCAAGATCAA 849
 |||||||
 DB 601 ACCGTGTCAGAGATCGAGACCTTCTGACAGCAATGGCACTGTGCTCCGATCAAGATCAA 660

OY 850 GAAGGATGAAAGATGGCGCTTACACCTCCCATGGTTCCACCCAGAAATGTCTCGGCTTC 909
 |||||||
 DB 661 GAAGGATGAAAGATGGCGCTTACACCTCCCATGGTTCCACCCAGAAATGTCTCGGCTTC 720

OY 910 AGCATTCGAAACCGCTCATCTATATAAGCGTGTGATCATGATGATGATGATGATGATGAT 969
 |||||||
 DB 721 AGCATTCGAAACCGCTCATCTATATAAGCGTGTGATCATGATGATGATGATGATGATGAT 780

OY 970 GAAGGCTCAGCAACCTGATGTCTGCCAATACCCAGAAAGCTTCCCTGAGGATGAGCTC 1029
 |||||||
 DB 781 GAAGGCTCAGCAACCTGATGTCTGCCAATACCCAGAAAGCTTCCCTGAGGATGAGCTC 840

OY 1030 ATGACGTGGCAGTTTGTGCTTCTGCAACCTGCGGGCCAGCGTCTCTTCTCAACTTC 1089

|||||
 DB 841 ATGACGTGGCAGTTTGTCTCTGCAACCTGCGGGCCAGCGTCTCTTCTCAACTTC 900

OY 1090 AACCTTCGAACTGAGAGAGAGAGAGGCGGTTGAATCTCATCTCCGGGCTCCACC 1149
 |||||||
 DB 901 AACCTTCGAACTGAGAGAGAGAGAGGCGGTTGAATCTCATCTCCGGGCTCCACC 960

OY 1150 ACCAACCOCGAGGTTCACAGCTGAGAGACAGAGCTGGGAACATGGCGGGAATTC 1209
 |||||||
 DB 961 ACCAACCOCGAGGTTCACAGCTGAGAGACAGAGCTGGGAACATGGCGGGAATTC 1020

OY 1210 AACCTTCCTGCAAGGCTGTGACCAAGATGCCCAAAAGTCCAGGATCTCCGGCTGAG 1269
 |||||||
 DB 1021 AACCTTCCTGCAAGGCTGTGACCAAGATGCCCAAAAGTCCAGGATCTCCGGCTGAG 1080

OY 1270 TTCCAGTTTGTGTCACATCCACAAATGAAGCAATTA 1310
 |||||||
 DB 1081 TTCCAGTTTGTGTCACATCCACAAATGAAGCACTGA 1121

RESULT 12
 AK026329 976 bp mRNA linear PRI 29-SEP-2000
 LOCUS Homo sapiens cDNA: FLJ22676 fls, clone HS110569.
 DEFINITION AK026329
 ACCESSION AK026329.1 GI:10439158
 VERSION AK026329.1
 KEYWORDS oligo capping; fls (full insert sequence).
 SOURCE Homo sapiens human small intestine cDNA to mRNA, clone_11b.HSI
 clone:HS110569.

ORGANISM

Homo sapiens
 Eukaryota; Chordata; Cranialata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE

Matanabe, K., Kumagai, A., Itakura, S., Yamazaki, M., Tashiro, H.,
 Ota, T., Suzuki, Y., Oba-yashi, M., Nishi, T., Shibahara, T., Tanaka, T.,
 Nakamura, Y., Isogai, T. and Sugano, S.
 NEDO human cDNA sequencing project

JOURNAL

AUTHORS

TITLE

Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,
 University of Tokyo, Laboratory of Genome Structure Analysis, Human
 Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,
 Japan (E-mail:cdna@leims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
 Fax:81-3-5449-5416)

COMMENT

NEDO human cDNA sequencing project supported by Ministry of
 International Trade and Industry of Japan; cDNA full insert
 sequencing: Research Association for Biotechnology; cDNA library
 construction, 5'- & 3'-end one pass sequencing: Department of
 Virology and Human Genome Center, Institute of Medical Science,
 University of Tokyo (partly supported by Science and Technology
 Agency).

FEATURES

source

1..976
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="HS110569"
 /tissue_type="human small intestine"
 /clone_11b="HS1"
 /note="cloning vector pME18SFL3"
 BASE COUNT 251 a 257 c 219 g 249 t
 ORIGIN

Query Match 15.6%; Score 959; DB 9; Length 976;
 Best Local Similarity 99.5%; Pred. No. 3.3e-191;
 Matches 962; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 4956 AAATGACTCATGTTATTAATGATTAATCACTTGAAGATGACCGGCTCAACTGTTCA 5015
 |||||||
 DB 1 AAATGACTCATGTTATTAATGATTAATCACTTGAAGATGACCGGCTCAACTGTTCA 60


```

OY 5016 CATGGCAATAATGAAGAATTTTATATGTCCTCTAAATGGCGTGTACTGCAGACCTCTT 5075
DB 61 CATGGCAATAATGAAGAATTTTATATGTCCTCTAAATGGCGTGTACTGCAGACCTCTT 120
OY 5076 GAACACTTTCCAGAGATAGATATTTAATGATGCGCTTGGGCTTCCCTATGGCACTT 5135
DB 121 GAACACTTTCCAGAGATAGATATTTAATGATGCGCTTGGGCTTCCCTATGGCACTT 180
OY 5136 TCCCTTTCTAAGAGTGTGTCTCTGCGCCAGTACCTTGGCGCTTGTAGCGGAGATCTGA 5195
DB 181 TCCCTTTCTAAGAGTGTGTCTCTGCGCCAGTACCTTGGCGCTTGTAGCGGAGATCTGA 240
OY 5196 CCTGCAATAAGGGCCAAAGAGGGGCTGGGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5255
DB 241 CCTGCAATAAGGGCCAAAGAGGGGCTGGGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 300
OY 5256 TGAATTCAGTGTGTGAGGCTTACCCCTCATTTCTGACATTTCCCACTCCCACTCC 5315
DB 301 TGAATTCAGTGTGTGAGGCTTACCCCTCATTTCTGACATTTCCCACTCCCACTCC 360
OY 5316 TTCCAAAGCAGAGTAGTGTGCTTCCCTGATTCACCCAAAGTGGGATTTGGCTTACGCT 5375
DB 361 TTCCAAAGCAGAGTAGTGTGCTTCCCTGATTCACCCAAAGTGGGATTTGGCTTACGCT 420
OY 5376 GGCTACTTGTCAACATCACCGACATCATGTTGGCTTGTCAAGGACACAGTGGCACTTTT 5435
DB 421 GGCTACTTGTCAACATCACCGACATCATGTTGGCTTGTCAAGGACACAGTGGCACTTTT 480
OY 5436 CCTTCACTGAGGGCTCAAAATCTCTGAGCAAGTGTGCTGCTCTCTGAGCAGATATTTCC 5495
DB 481 CCTTCACTGAGGGCTCAAAATCTCTGAGCAAGTGTGCTGCTGCTGAGCAGATATTTCC 540
OY 5496 TGGAGTGTGCTCTCAATGAAAGGGGCCACCTGTAGAGAACCCCTGCTTTTCTTAAAGC 5555
DB 541 TGGAGTGTGCTCTCAATGAAAGGGGCCACCTGTAGAGAACCCCTGCTTTTCTTAAAGC 600
OY 5556 CCAGGCCCACTTACATTAATAACATTTCAAGGCTCTCTGAAAGCAAGTGGCACTTTGTT 5615
DB 601 CCAGGCCCACTTACATTAATAACATTTCAAGGCTCTCTGAAAGCAAGTGGCACTTTGTT 660
OY 5616 GAAGCTTACTGATGCGACGCCCACTGCTCATTCACAGTGTGCTCTGAGCAGAGAA 5675
DB 661 GAAGCTTACTGATGCGACGCCCACTGCTCATTCACAGTGTGCTCTGAGCAGAGAA 720
OY 5676 GCCAGGCGATGCAAGCACTGCTCTTAATGCTGTGTGCTATTTGCAAGAGGAAAGTCTC 5735
DB 721 GCCAGGCGATGCAAGCACTGCTCTTAATGCTGTGTGCTATTTGCAAGAGGAAAGTCTC 780
OY 5736 AAGGAAGAGTCAACTGGGCAAGCAAGCAAGCCGACATGGGCTTGGTAAAGTTAGC 5795
DB 781 AAGGAAGAGTCAACTGGGCAAGCAAGCAAGCCGACATGGGCTTGGTAAAGTTAGC 840
OY 5796 AGACTGTGTGTGTGATGCTGAGTGTCTCACTGGAATAATTTATTTCAATTCAGATAC 5855
DB 841 AGACTGTGTGTGTGATGCTGAGTGTCTCACTGGAATAATTTATTTCAATTCAGATAC 900
OY 5856 TTTTATGATGCGATTTTATTTATTTCTCTGTGCTTTAAATAACAATGTACCAAAAAACA 5915
DB 901 TTTTATGATGCGATTTTATTTATTTCTCTGTGCTTTAAATAACAATGTACCAAAAAACA 960
OY 5916 GTATCA 5922
DB 961 AAAAAA 967

```

```

RESULT 13
LOCUS BC011340 3843 bp mRNA linear ROD 07-AUG-2002
DEFINITION Mus musculus, clone IMAGE:3709937, mRNA, partial cds.
ACCESSION BC011340
VERSION BC011340.1 GI:15030169
KEYWORDS
SOURCE house mouse.

```

```

ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 3843)
JOURNAL Strausberg, R.
DIRECT SUBMISSION Submitted (25-JUL-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Offices, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cghabs-remail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Gunnar, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov
Series: IRAC Plate: 17 Row: 1 Column: 17
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis.
FEATURES
source location/Qualifiers
1..3843
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="FVB/N"
/clone="IMAGE:3709937"
/tissue_type="Mammary tumor, C3(1)-Tag model. Infiltrating
ductal carcinoma. 5 month old virgin mouse."
/clone_lib="NCI CGAP Mam6"
/lab_host="DH10B"
/note="Vector: pCMV-Sport6"
<1..1155
/codon_start=1
/product="Unknown (protein for IMAGE:3709937)."
/protein_id="AA11340.1"
/db_xref="GI:15030170"
/translation="LLVPKDKLSLMLVPGOKLQOHTOERPCNTSFYHYVSTTPGODL
YRSPSGSGSIRKIOYKONSSVTLRAYAPSPQDEYSKGLIYSTPYPEEGEITFTPT
DPRNKVYLSPHMDRLPALSSVSNISPSNOVACTYLKRSGLACSGAPMIIQ
EEOSSRAEELFSLSEEVLPKPSFHHHSFWNISCSPMNKQDLLEWVLTLPRTYDLA
VVIAGAGGALLFLALVLIICFKKKKKYKDPAYGIVGNVNTQMPQTPKPKGRKD
NDSHYAVIEDPMYVGHLLQDSGGSSIQEVEVTPRFQSPMDGCPPTPLPSRPTA
KFTAEELASSPSESESEPTFESHPKSGIIVRENDIPILHTQGVETEE"
BASE COUNT 1015 a 943 c 906 g 979 t
ORIGIN
Query Match 13.3%; Score 821.2; DB 10; Length 3843;
Best Local Similarity 74.1%; Pred. No. 3e-162;
Matches 1183; Conservative 0; Mismatches 378; Indels 35; Gaps 10;
OY 1630 CTGCTGGTGGCCCAAGAGAGCTCAGCGTGTCTGTGTCACCCAGAGAGCTGCAGCAG 1689
DB 1 CTACTGTGGCCCAAGAGAGCTCAGCGTGTCTGTGTCACCCAGAGAGCTGCAGCAG 60
OY 1690 CATACACAGAGAGAGCCCTGCAACACAGCTTACGCTGTGGCAGTGCATACCC 1749
DB 61 CATACACAGAGAGAGCCCTTAAACACAGCTTGGGAGTGCATGACAGCAGCAGCCCT 120
OY 1750 AGCCAGAGCTGTACTTGGCTCTTGTGCGCGGAGAGCTCTATCAACAGCAGATCAGGTG 1809
DB 121 GCCCAGAGAGCTGTACTTGGCTCTTGTGCTGAGAGAGCTCATGCAAGAGATCCAGGTG 180

```


QY	1810	AAGCAGAAACATCTGGGAGACCTTTCGACACTTTGGCCCGACGTTCCAAACAAGAGCCATC	1869
Db	181	AAGCAAAATAGCTGGGTGACACTGACAGCCCTATGCTCCACGCTTTCACAGAAAGTCTCC	240
QY	1870	AGGCAGGGGTCTGACGGGTGCTCTTATACCTTATTTCCAAAGAGAAAGCGCTTTTCAGGGT	1939
Db	241	AAGCAAGGCCCTGATAGTGTCTCTATACACCGGTATTTCCAAAGAGAAAGCGCATCTTCACGGTG	300
QY	1930	ACCCTTGACACAAAAGCAAGGTCTACCTGAGAACCCCACTGGAGCCGGGGCCTGGCA	1988
Db	301	ACCCCAACACAAAAAACAAGGTCTACCTAAGATCACCCACACTGGAGACCTGTGGCCTGGCT	360
QY	1990	TCCCTACGCTGTGTCTGCTGGAACTACAGGGTCCCAAGAGACAGAGTGGCCGTGGTACT	2049
Db	361	GCCTCTCTCTAGTGTCTTGGAACTACAGTGTGCTGAGAACCAAGTGGCTGTCTGACCC	420
QY	2050	TTCTTTAAGAGAGCGAGACGGCGCTGTGTCTGCGACAGACGGGGCGGCATTCATGATCATCCAG	2109
Db	421	GTCTTGAAGAAGGGTCTGTGGCTTGGCCTGTGCAGTCAAGGGCGTGATTCATGATCATCCAG	480
QY	2110	GAGCAGCGAGACCGGGGCTGAGAGATCTTCAAGCTTGAGAGAGAGATGTGCTCCCAAGCCA	2169
Db	481	GAGCAGCAGATCCGGGCGAGAGAGATCTTCAAGCTTGGAGAGAGAAAGTCCGTCCCTAAGCCA	540
QY	2170	AGCTTCCACCATCACAGCTCTTGGGTCAACATCTCTAATGAGGCCACAGCGGGCAAG	2229
Db	541	AGTTTCCACCATCACAGCTCTTGGGTAACTCTTCCAAATTTGAGGCCCATGATATGGCAG	600
QY	2230	CAGCTAGACCTGCTCTCTGCTGGTGACACTTACCCCAAGAGACTGTGGACTTGTCACTGATC	2289
Db	601	CAGTCTAATTTGCTCTTCTGGGTGACTCTTACCCGAGAGACTGTAGATTTGGCTG---TC	657
QY	2290	CTCATGCGACGGGTGGGAGGTGAGTCTTACTGTCTGTGCCCCGTCTATCATTTTGC	2349
Db	658	GTCAATTGGTGCAGAGAGAGGTGGAGCTCTGTGCTGTGTGGCCCTGGTACTCATCTATGCG	717
QY	2350	TGTGTGAAAAAAGAAAAAAGAACAAACAAAGAGGCCCGCTGTGGGTATCTCAAAATGGC	2409
Db	718	TTGCTGAAAAAAGAAAG---AAGAAAGGTGATTAAGAGGTCTGTGTGGGTATCTATATATGGC	774
QY	2410	AACATCATATCTAGTGAATGCCGAGAGCAGCCAAAAAATTTCAAAAAAGGGCGAAAGACAT	2469
Db	775	AATGTCAATACCGAGATGCC---CCAGACCCAAAAGTTCCGAAAAGGGCGAAAAGACAT	831
QY	2470	GACTCCCATGTGTATGCACTATCGAGAGACACCATGTGTATGTGGCAATCTGCTACAGGAT	2539
Db	832	GACTCCCATGTGTATGCTGTCTATCGAAGAACACCATGTTATGTGGCACTTGTCTGACGAGAC	891
QY	2530	TTCAGGGGGTCTCTCTGAGAGCCAGAGAGTGGAGACACTACCGGCCGTTCCAGGGCACATG	2589
Db	892	TTCGGGTGATCTTCTATCTACCAACCAAGAGTGGAGACCTTACCGGCCCTTCCAGGGCCCCATG	951
QY	2590	GGGGTCTGTCTCTCTCTCCACCCACCCACCATATGTCTCCAGGGGCCCAACTGCAAAAGTTGGCC	2649
Db	952	GGAACTGTGCGCTCTCCACCCCAACCCCTCTATTTCTCAGGAGCCCAACTGTCAAAAGTTCACT	10111
QY	2650	ACTGAGAGCCACCTCTCTGCTCCCTCCCTCTGAGTGTGAGAGTAAACCGTACACTTTCC	2709
Db	1012	GCAGAGAGAGCTGGCTCTCTATATAGCCCTCTCCGAGTGTGAGAGTAACTTACCTTTCTCG	10711
QY	2710	CATCTCCCAACAATGGGATGTGAACAGACAAAGACACAGACATTTCCCTTACTGAACACTAG	2769
Db	1072	CATCTCCCAACAAGGGGAGATTTGGTGTATGGAGAGTCAATCCCTCTACTCCATCCCAA	11311
QY	2770	GAGCCCATGGAGCGACAGAAATTAATTT-GATTCATTTCCAGAGCGTTTGTGTGATTTCAATA	2828
Db	1132	GGACCAATGGAGACTGAGAGTAACTTGTGGCCCATTTCCAAAACATTTGTGTGATGATCG	11911
QY	2829	AAGCAGGGCACTGAGACCCCTCTCCCTGTGTCTTAACCAAAATCTCTAAAGAGAGATTT	2888
Db	1192	GCACCAACAATTTGAACATTAABAACCTG---TCCATCTAAGAAATCTCAGAGGAAAGCCCAATC	1250
QY	2889	ATACAGAGAGAACGACGAGAGTTTTTCTGTGACACCCGCCAACCTTCACATTTGCTCATGTGA	2948

Db	1251	AGCTAGAAAGGA---GAGTAGCTTTTGGCAAGGCGCCATCAATCTTGATCTTCTGGTGGG	1307
Qy	2949	CTCATTTCTAAGGGCAACAGCTTGAATAATGATGANTTCCATCTGGATACAGTCATGACAG	3008
Db	1308	CTGCTT-----CCAGTAGTGGGCAAAATCATGACTTCTATGAGATGCCATCCG-AAAG	1359
Qy	3009	CTCATNGCTCTCTCAATTTTATGGCTGGGCGTTAGCCAGCGTCTAATGAGAGAGAGAGGC	3068
Db	1360	CTTCAGGCGCTTCTCGAATCTCGGGTTG--GTGGGGCTGGGCACAGGAGACGAGAGAGAC	1417
Qy	3069	CTGAGTCACTAGCATATGAGGTTGCACACAAGCCCTGATTCAGAGTGTAAACAGAGGCTT	3128
Db	1418	ACGGTCCACCAAGACAAGACTACAATAGAGCCCTGGAGTCTG-----GATTCGCC	1466
Qy	3129	GCCTCTTTCACAGACACAGTTCCAAATCCAAAGAGGCTTACCTGAGGTCCTACTACT	3188
Db	1467	GATGACACAGCAGAGACTTGGATGATTCGCCAGGAGCTCCACTTGGAGGCTCTGCTCTCACT	1526
Qy	3189	GGGGTCCCAAGATGAAACGACAAATGCGCTTTT	3224
Db	1527	GGAGACCTATGATGAGAAATAGCCAAATGTGCTTTTTT	1562

LOCUS	BC034137	1999 bp	mRNA	linear	ROD 07-AUG-2002
DEFINITION	Mus musculus, similar to hypothetical protein FLJ22969, clone MGC:32174 IMAGE:5004246, mRNA, complete cds.				
ACCESSION	BC034137				
VERSION	BC034137.1	GI:21707874			
KEYWORDS	MGC.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	1 (bases 1 to 1999)				
TITLE	Strausberg, R.				
JOURNAL	Direct Submission				
REMARK	Submitted (02-JUL-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA				
COMMENT	NIH-MGC Project URL: http://mgc.ncl.nih.gov Contact: MGC help desk Email: cgapbs-remail.nih.gov Tissue Procurement: Gilbert Smith, Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305 Web site: http://www.shgc.stanford.edu Contact: (Dickson, Mark) mcdep@llnl.stanford.edu Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.				
FEATURES	<p>Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IRAK Plate: 44 Row: P Column: 21 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.</p> <p>Location/Qualifiers</p> <p>1..1999</p> <p>/organism="Mus musculus"</p> <p>/db_xref="taxon:10090"</p> <p>/map="FVB/N-3"</p> <p>/clone="MGC:32174 IMAGE:5004246"</p> <p>/tissue_type="Mammary tumor, MMTV-LTR/INT3 model, 5 month old mouse, Taken by biopsy."</p> <p>/clone_idb="NCI CGAP_Mam2"</p> <p>/lab_host="DH10B"</p> <p>/note="Vector: pCMV-SPORT6"</p>				
SOURCE					

CDs 99.1061
/codon_start=1
/product="similar to hypothetical protein FLJ22969"
/protein_id="AAH34137.1"
/db_xref="GI:21707875"
/db_xref="LocustID:201643"
/translation="MOEYKMAHLPMFHRNVSFSLNRSRIKLTIESVEGEG
SATLSANYPGPEDELMTOFVPAHRAVSFLNFSNCEKREVEYIIGST
TNEVFRLEDKPGNMAGNENISLOGCCODNASPILRQEVLYORPDESNKTYM
DISREPTSLTEPRPVKHNRFVPGCEVLESRCSTNTVLTASIRKISFLCDLT
RLMVERKLSMPLKPLTSLISLPSLSDISPHPLPOLCAHARTHTHTHTHYCK
GSEDLNMFSEFCHVDSDQVOALQSLMCHPLSLHLSAQ"

BASE COUNT 494 a 552 c 469 g 484 t

Query Match 10.1%; Score 619.4; DB 10; Length 1999;
Best Local Similarity 87.1%; Pred. No. 9e-120;
Matches 680; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

OY 749 CAGACGAGTCACTCACTCCATCAGGCGGCAATGCATGCACCGGTGTCAGATGGAA 808
DB 4 CAGACGGGTGACTCACTCCATCAGTGGCCATTAATGATGCCTGAGATGGGA 63
OY CCTTTCGACAAATGGCACTGTGTCCCGATCAAGATGCAAGAAGAGTGAATGGCCT 868
DB CCTTTCGACAAATGGCACTGTGTCCCGATCAAGATGCAAGAAGAGTGAATGGCCT 123
OY 869 TACACTCCCATGCTTCCACCCCAAGAAATGTCCTCCGCTTCAGCAATTCGAAACCTCAT 928
DB 124 TACACTCCCATGCTTCCACCCCAAGAAATGTCCTCCGCTTCAGCAATTCGAAACCTCAT 183
OY 929 CATTAAGCTGTGATCATCATGATGATGTTGAGGGTGAAGGCTGCAACCTCA 988
DB 184 CATTAAGCTGTGATCATCATGATGATGTTGAGGGTGAAGGCTGCAACCTCA 243
OY 989 TGTCTCCCACTACCCAGAGGCTTCCAGAGATGATGATGATGATGATGATGATG 1048
DB 244 TGTCTCCCACTACCCAGAGGCTTCCAGAGATGATGATGATGATGATGATGATG 303
OY 1049 TTCCTGACAACTGCGGCGGCAAGCTCTCTCTCAACTTCAACTTCTCTCAACTG 1108
DB 304 TTCCTGACAACTGCGGCGGCAAGCTCTCTCTCAACTTCAACTTCTCTCAACTG 363
OY 1109 GGAAGAGAGGCGGTTGAATACATCCCGGCTCCACCAACCCCGAGGTGTCA 1168
DB 364 GGAAGAGAGGCGGTTGAATACATCCCGGCTCCACCAACCCCGAGGTGTCA 423
OY 1169 AGCTGAGAGCAAGCAGCTGAGAACATGCGGGAACTTCAACTTCTCTCAAGGCT 1228
DB 424 AGCTGAGAGCAAGCAGCTGAGAACATGCGGGAACTTCAACTTCTCTCAAGGCT 483
OY 1229 GTGACCAAGATGCCCAAGATCTCCGCTGCAATTCGAAATTTGGTCAAC 1288
DB 484 GTGACCAAGATGCCCAAGATCTCCGCTGCAATTCGAAATTTGGTCAAC 543
OY 1289 ATCCACAATAAGAAACATTAATCTAGCTGTGATGATGATGATGATGATGATG 1348
DB 544 ATCCACAATAAGAAACATTAATCTAGCTGTGATGATGATGATGATGATGATG 603
OY 1349 CACTCACTGACGAGCGGCGGCTCAACAGAGCGCAAGTGTGCGGTGGTGGT 1408
DB 604 CACTCACTGACGAGCGGCGGCTCAACAGAGCGCAAGTGTGCGGTGGTGGTGGT 663
OY 1409 TGTGTCTAGAAATCTGCACTGATGATGATGATGATGATGATGATGATGATG 1468
DB 664 TGTGTCTAGAAATCTGCACTGATGATGATGATGATGATGATGATGATGATG 723
OY 1469 AATCTCTCTCTTGTGATGATGATGATGATGATGATGATGATGATGATGATG 1528
DB 724 AATCTCTCTCTTGTGATGATGATGATGATGATGATGATGATGATGATGATG 783
OY 1529 G 1529

DB 784 G 784

RESULT 15
AX192795 544 bp DNA linear PAT 15-AUG-2001
LOCUS AX192795
DEFINITION Sequence 362 from Patent WO0149716.
ACCESSION AX192795
VERSION AX192795.1 GI:15210751
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1 (bases 1 to 544)
Xu, J., Lodes, M. J., Secrist, H., Benson, D. R., Meagher, M. J.,
Stolk, J. A., King, G. E., Wang, T., and Jiang, Y.
Compounds for immunotherapy and diagnosis of colon cancer and
methods for their use
Patent: WO 0149716-A 362 12-JUL-2001;
CORIXA CORPORATION (US)
LOCATION/Qualifiers
1. 544
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 138 a 123 c 113 g 170 t

Query Match 8.8%; Score 542.4; DB 6; Length 544;
Best Local Similarity 99.8%; Pred. No. 1.5e-103;
Matches 543; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3068 CCGAGTCACTACATGATGATGATGATGATGATGATGATGATGATGATGATG 3127
DB 1 CCGAGTCACTACATGATGATGATGATGATGATGATGATGATGATGATGATG 60
OY 3128 TGGCCTCTTCCAGCAACAGTTCATTCACAGAGGCTGAGGCTCTACTCTCAC 3187
DB 61 TGGCCTCTTCCAGCAACAGTTCATTCACAGAGGCTGAGGCTCTACTCTCAC 120
OY 3188 TGGGCTCCCGAGATGAAAAGCAATGTCCTTTTATTTATTTATTTATTTATTT 3247
DB 121 TGGGCTCCCGAGATGAAAAGCAATGTCCTTTTATTTATTTATTTATTTATTT 180
OY 3248 GTGTATTATTAAGATCAATATTAACACCTAGCTCTTTTCACTGATGATTA 3307
DB 181 GTGTATTATTAAGATCAATATTAACACCTAGCTCTTTTCACTGATGATTA 240
OY 3308 CTCTACTACGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATG 3367
DB 241 CTCTACTACGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATG 300
OY 3368 TGGCTGCCCGCAGGTGGTGGATATTTACATCTGTCCAAACGAAAGATGTGT 3427
DB 301 TGGCTGCCCGCAGGTGGTGGATATTTACATCTGTCCAAACGAAAGATGTGT 360
OY 3428 TGTTTGAGCAGATGACACATATCTGCTTTGATTAAGAGACTTCTGATTTCT 3487
DB 361 TGTTTGAGCAGATGACACATATCTGCTTTGATTAAGAGACTTCTGATTTCT 420
OY 3488 GTTCGCTGATATCCCATTTGGAATTCATCTGAATCCCATTTGCTTAAGCTCA 3547
DB 421 GTTCGCTGATATCCCATTTGGAATTCATCTGAATCCCATTTGCTTAAGCTCA 480
OY 3548 ATAAAGAAATTTCTCAATTTCTCAATGTCGCTTCTCAAGTGCAGCAATCTT 3607
DB 481 ATAAAGAAATTTCTCAATTTCTCAATGTCGCTTCTCAAGTGCAGCAATCTT 540
OY 3608 ATTT 3611
DB 541 ATTT 544

Search completed: November 11, 2002, 09:32:45
Job time : 16209 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 11, 2002, 11:31:42 ; Search time 22 Seconds

(without alignments)
1118.071 Million cell updates/sec

Title: US-09-899-569A-4

Perfect score: 4394

Sequence: 1 MAGLNGCYSIALGLVLLGA.....SSKDDIDPLNTQEPMEPAE 836

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database :

Issued Patents, AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCITUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	117	2.7	2476	2	US-08-276-967-2
2	116	2.6	1589	3	US-08-755-587-189
3	110.5	2.5	1719	2	US-08-459-568-4
4	110.5	2.5	1719	2	US-08-399-411-4
5	110.5	2.5	1719	3	US-08-516-859A-4
6	110.5	2.5	1719	4	US-09-586-472-4
7	110.5	2.5	1719	4	US-09-528-706-4
8	106.5	2.4	2343	4	US-09-324-867-2
9	106	2.4	670	2	US-08-473-750-11
10	106	2.4	670	2	US-08-477-325-11
11	105.5	2.4	683	4	US-09-213-295D-1
12	104	2.4	449	2	US-08-839-008-2
13	103.5	2.4	727	5	US-08-475-844-9
14	103.5	2.4	727	5	PCR-US95-08429-9
15	103	2.3	449	2	US-08-839-008-9
16	102	2.3	2183	3	US-08-746-111-5
17	101.5	2.3	535	4	US-09-117-860-24
18	100.5	2.3	468	4	US-08-839-008-7
19	100.5	2.3	468	4	US-09-032-523-8
20	98.5	2.2	567	1	US-08-007-2828-2
21	98.5	2.2	1481	2	US-08-616-844-40
22	98.5	2.2	1481	2	US-08-599-654-40
23	98.5	2.2	1481	3	US-08-944-868A-40
24	98.5	2.2	1481	3	US-08-944-423A-40
25	98.5	2.2	1481	3	US-08-944-496-40
26	97.5	2.2	318	4	US-09-068-051A-32
27	97.5	2.2	1785	4	US-09-341-587-3

28	96.5	2.2	1332	2	US-08-971-244-2	Sequence 2, Appl1
29	96.5	2.2	1332	2	US-09-286-891-2	Sequence 2, Appl1
30	96	2.2	401	2	US-08-839-008-5	Sequence 5, Appl1
31	95.5	2.2	424	1	US-08-453-472-8	Sequence 8, Appl1
32	95.5	2.2	424	1	US-08-038-948-2	Sequence 2, Appl1
33	95.5	2.2	424	1	US-08-038-948-5	Sequence 5, Appl1
34	95.5	2.2	424	1	US-08-453-952-8	Sequence 8, Appl1
35	95.5	2.2	424	2	US-08-862-903-8	Sequence 8, Appl1
36	95.5	2.2	703	4	US-09-433-043B-124	Sequence 124, App
37	95.5	2.2	1706	2	US-08-459-568-2	Sequence 2, Appl1
38	95.5	2.2	1706	2	US-08-399-411-2	Sequence 2, Appl1
39	95.5	2.2	1706	3	US-08-516-859A-2	Sequence 2, Appl1
40	95.5	2.2	1706	4	US-09-586-472-2	Sequence 2, Appl1
41	95.5	2.2	1706	4	US-09-528-706-2	Sequence 2, Appl1
42	95	2.2	749	4	US-09-562-737-93	Sequence 93, Appl
43	95	2.2	4302	3	US-08-658-136-5	Sequence 5, Appl1
44	95	2.2	4302	4	US-09-052-469-8	Sequence 8, Appl1
45	95	2.2	4303	2	US-08-460-751-2	Sequence 2, Appl1

ALIGNMENTS

```
RESULT 1
US-08-276-967-2
Sequence 2, Application US/08276967
Patent No. 5851817
GENERAL INFORMATION:
APPLICANT: Hardy, Daniel M.
TITLE OF INVENTION: Species-Specific Egg-Binding Proteins of
TITLE OR INVENTION: Sperm
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Arnold, White & Durkee
STREET: P. O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/276,967
FILING DATE: Submitted Herewith
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Kitchell, Barbara S.
REGISTRATION NUMBER: 33,928
REFERENCE/DOCKET NUMBER: UTSD:418\KIT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 713-787-1400
TELEFAX: 713-789-2679
TELEX: 79-0924
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2476 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-276-967-2

Query Match      2.7%  Score 117;  DB 2;  Length 2476;
Best Local Similarity 21.6%;  Pred. No. 0.19;
Matches 94;  Conservative 53;  Mismatches 140;  Indels 148;  Gaps 23;

OY 484 VASAIISODLYFSPFGSIRKIQVKON-----ISVTLRFAP- 522
DB 85 VLGISIRKHILFSGQ--PGRSMQPVSVNYTSQGIQFTLVGVRKIRPEPAVDAISIAIC 142
```

```

QY 523 --SFQO-----EAS-----RGLTVSFIPIFKEGVETVPDITKSK-VYL 559
Db 143 EESFPQCFEDNAHPCFDWQASODGGYWRGKNKFTL---QAPAGPGISLNGEYIFL 199
QY 560 RTPMWDGLPSLTISVSNISVPRQOVACLTFFKERSGVCOGTGAFTMIOQRRAEIEF 619
Db 200 ETDKFSQAGOSFRLVSPFCAP-AVICVFETHMGL-----GQGTKLRL 245
QY 620 -----SLDEDVLPK-PSFHHSFWVNISNCSPTSGKO-LDLFSVLTLPRTVDLTV 668
Db 246 GSPAGSPSSLMERVGQSPK-----WLNSTVPIPSGHQOPLMILFE----- 287
QY 669 ILAAGVGVLISALDILICVKKKKKTKNGKPAVGIYNGNINTEMPROPKFKQGRKD 728
Db 288 ---AVRGNTAFVVALGFLV-----INHGTCRGSPSETSVSTKRPVAPTE---KPT 331
QY 729 NDSHVYV-IEDTMVYGHLLDSSGSLQRPVDIYRPFQGMGVCPPSPITICRAPATK 787
Db 332 VPSEIYTIPTKEKPVH-----MEKPIVHTKRP---TVPTKEPTIPEKSTVPTK 378
QY 788 -----LATEBP-----PRSPSESESE-----YTFSPNNGDVSSKDTDIP- 824
Db 379 PTVKEPTLPREGPTVAERTTPREGPAVPKGPVLTETMPTSHTEKSTVHTKEPILPT 438
QY 825 ---LNTQEPMEPAE 836
Db 439 GKSTIPTKEKPVPTK 453

```

RESULT 2

US-08-755-587-189

; Sequence 189, Application US/08755587

; Patent No. 6045997

; GENERAL INFORMATION:

; APPLICANT: Futreal, Phillip A

; APPLICANT: Wooster, Richard F

; APPLICANT: Asmworth, Alan

; APPLICANT: Stratton, Michael R

; TITLE OF INVENTION: Materials and methods relating to the

; TITLE OF INVENTION: Identification and sequencing of the BRCA2 cancer

; TITLE OF INVENTION: susceptibility gene and uses thereof.

; NUMBER OF SEQUENCES: 222

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Bell Seltzer Park & Gibson

; STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107

; CITY: Raleigh

; STATE: NC

; COUNTRY: USA

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/755,587

; FILING DATE: 25-NOV-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: GB 9523959.6

; FILING DATE: 23-NOV-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: GB 9525555.0

; FILING DATE: 14-DEC-1995

; APPLICATION DATA:

; APPLICATION NUMBER: GB 9617961.9

; FILING DATE: 28-AUG-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Kenneth D Sidley

; REGISTRATION NUMBER: 31,665

; REFERENCE/DOCKET NUMBER: 5405-135

; INFORMATION FOR SEQ ID NO: 189:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1589 amino acids

```

; TYPE: amino acid
; TOPOLOGY: linear
; US-08-755-587-189
Query Match 2.6%; Score 116; DB 3; Length 1589;
Best Local Similarity 17.7%; Pred. No. 0.11;
Matches 134; Conservative 105; Mismatches 233; Indels 286; Gaps 34;

```

```

QY 198 LPWFHPRNVSGF-----SIARRSSIKRLCILESYEEGSAATLMSANVPGEFDEL 249
Db 558 LPNINENEPFGFCALTKLISVNEALRKAMKLFSDIENSEPESAKVG---PRFSS-- 612
QY 250 MTQWVPAHLRASVSFLNRLNSCERKEER-----VEYIPGST-----INPEYFKL 297
Db 613 -----AAHDSVASVFYKIKQNTKESDKSSKQVTLQNNIEMTTCIFVGRNPEKTYK 664
QY 298 EDKQGNMAGNFNLISLGGCDQDQSPILRLQFVLQVQHQNESNKLYV---VDLSNER 353
Db 665 NTKHEDSYTSSQRNNLNSGSMSS-----TSGPVYIHKGSDL--- 703
QY 354 AMSLTIEPRPVKOSRKFPVPCFVLESRTCSNLTLSGSKHKISFLCDDLTRLMMVER 413
Db 704 -----PADQSK-----CPESCQYARENTQ-IKENIS---DLTCLIMAE 742
QY 414 TISCTDHRICORKSYSLQVPSDILHPEYLHDSW----- 448
Db 743 T-----CMKSSDKKQLPSDKMEONIKEFNISFQASGKNTRVSKESLNSVNIENRE 794
QY 449 --KLVPKDRL-SLVYPAQKLOQT--HEKPCNTSPYLVASIPRODLYFGSFCPGS 503
Db 795 TDELTVISDSLSNKLILGINKDKMHTSCHK-----ALSTKKYFEH--PIYV 840
QY 504 IKQIQVKNISVTLRTPAPSPQOEASRQGLTVSE-----IPYKPE--EGVFTVPTDTK 554
Db 841 VSQLPAGQH-----PEYETESTKEPTLLSPHTASGKKVKLMQSLDKVKKLFPDETQ 891
QY 555 SKYLLR-TPMWDKGLPSLTISVSNISVPRQOVACLTFFKERSGVCOGTGAFTMIOQRT 613
Db 892 ---YVRKTASFQSGSKPL-----KDSKELTLAYEK-----IEVTAS 925
QY 614 RAERI---FSLDEDVLPKPSFHHSFWVNISNCSPTSGK-QLDLFSVLTLPRTVDLTV 669
Db 926 KCEMKNFVSKETEMLQONIHMYROTENLKTSGTSSKVOENENNENVR----- 978
QY 670 ILAAGVGVLISALGILICVKKKKKTKNGKPAVGIYNGNINTEMPROPKFKQGRK-- 727
Db 979 -----ICICQSSYPVEDSALAYTEDSKRTCYRE--SSLSKGRKWL 1019
QY 728 -----DNDSHVYAVI-----EDTM 741
Db 1020 REQDKLGTNTIKIECKEHTEDFAGNASYEHSLVITRTETIDNHNVSNOVSTLLSDPNV 1079
QY 742 VYGHLLDSS-----SGSFLOPEVDIYRPFQGMGVCPPSPITICRAPAKLATE 791
Db 1080 CHSTLSQSSFCCHCDMDHNDGIFLAKNIDS-----DVPDPM-----KNAEGNTI 1123
QY 792 EPPRSPSESESEPTFSH-----PNNGDVSSKDTDI 823
Db 1124 SPVRSATKERNLHPQTLINCYCKLETNTSPHANNDAVI 1161

```

RESULT 3

US-08-459-568-4

; Sequence 4, Application US/08459568

; Patent No. 5811304

; GENERAL INFORMATION:

; APPLICANT: Huang, Shi

; TITLE OF INVENTION: Retinoblastoma Protein - Interacting

; TITLE OF INVENTION: Zinc Finger Proteins

; NUMBER OF SEQUENCES: 93

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Campbell and Flores

; STREET: 4370 La Jolla Village Drive, Suite 700

```

CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,568
FILING DATE: 02-JUN-1995
CLASSIFICATION: 514
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/399,411
FILING DATE: 06-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 1264
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1719 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-459-568-4

```

Query Match 2.5%; Score 110.5; DB 2; Length 1719;

Best Local Similarity 19.1%; Pred. No. 0.46; Indels 333; Gaps 42;

Matches 169; Conservative 96; Mismatches 289; Indels 333; Gaps 42;

```

QY 78 FTSCSPENHFV-----IEIQKNDICMSGCPREGEVQLQSTSLPLTNTFTMDVKAHK 133
DB 358 FMFPCOCHCKERTTKQGLERHMH-----ISTVNAF-----KCK 393
QY 134 STGLELOFSIPRLQ-----IGPESCPDGV-----HSISGRIDATVVRIGT 176
DB 394 YCGKACGTQINRRRHERHAGLRKRSQTLQPSDLADGKASGENVASKDDSPSLGP 453
QY 177 FC-----SNGTVSRIRKQEG-----VKMALHLWFPNRVNSGF 209
DB 454 DCLIMSEKASOPTINSSVEEYKVELHPCKYCKKVFCTHTNMRRHORVHERHLIPK 513
QY 210 SIANRSIRKLCITESFEEGEGSATLMSANYPRGEPDE-----LMTQGFVPAHLRA 262
DB 514 GVRKRGLE-----EPQRAEQAQATONVYVSTPEEEGEADVTIMD----- 557
QY 263 SVSFLNLSNCKERKEERYVIRG-----STTNPVEFKLEDKQPNMAGNENLSLQGC 316
DB 558 -----ISSNIS-----ENLNYIDGKIQTNNNTSNCDIVEME----- 589
QY 317 DQDQASGILRLQFOVLYQHPQN-ESKTIYV-DLSNERAMSLTIERPRYKOSKFPYGC 374
DB 590 SASADLYGINCLTPVIVETQNIKTQVPVTEDELREPUGSTNSE---AKKRRTASPPA 646
QY 375 FVCLERT-----CSSMLTSGSKHISFCLCDLTLMMNVKTICTDHRVQOK 426
DB 647 LPIKATETDPMVPSGSLPLSISTTEVSP-----HKESV-----YLSK 690
QY 427 SYSLQVPSDILHVELHDSWKLVPKRLSLVLPQAOKLQOHTHEKPCNTSESYLVA 486
DB 691 LKOLLQTODKLTPPAGISATEIAKLP---VCVSAPASML-----PYTSSRFKRTS 729
QY 487 AITSQDLFFSGFGGSIKIQYKQINISVTLRFA-PSFOQA-SRQGLTVSFTPYKKE 544
DB 740 SPSSSPH-----SPALRDFGKPSDKAAWTDAGLT----- 770
QY 545 GVTVPDTSKYVLRTPNMDRGLPSLTYSWMS--VPRDQVACLTFEFGSGVQQTG 602

```

```

DB 771 -----SKKSLESHS-----DSPAMLSGRDERETVSPCFDEYKMSKEWTAS 813
QY 603 RAEMIIOE-----RTRAEIRSLD--EDVLEKPSFHHHSFTVINSNCSPTSGKQ 650
DB 814 SAFSSVNOQOPLDLSSGVKOKAKBGTQTPVQWESVL-----DLVHKHKSDSGKE 865
QY 651 LDLLFSVTLPRIVDVLVILAAVGGVLLLSALGLICCVKKKKTKTN---KGPAVGIY 707
DB 866 FKSHSVQPT-----CSAVKKRKTMTMLQKVLINEX 897
QY 708 NG-NINTEMPR-----QPKFKQGRKNDSHYAVIEDVMYGHLLQDSSGSLQPEVD 760
DB 898 NGIDLPEVNPADGRSPSPCKSLQAQDPD-----LGPGSG-FPAATVE 940
QY 761 TYRPGCTGVCPPSP---PTICSR-----APPAKLATEEPP----- 794
DB 941 S-----TPDVCPSSPALQTPSLSSGQLPPLLPDTPDSSPPCPVLTVAATPPPLLPV 994
QY 795 PRSPSESEPTFTFHPNNGDVSKQTDIPLN-----TQEPMEPAE 836
DB 995 PLPAPSSASPHPCPSPLN--ATAQSPDPLTSLFTVSSPSPPLPVE 1039

```

RESULT 4

US-08-399-411-4

; Sequence 4, Application US/08399411

; Patent No. 5831008

; GENERAL INFORMATION:

; APPLICANT: Huang, Shi

; TITLE OF INVENTION: Retinoblastoma Protein - Interacting

; TITLE OF INVENTION: Zinc Finger Proteins

; NUMBER OF SEQUENCES: 93

; CORRESPONDENCE ADDRESS:

; STREET: 4370 La Jolla Village Drive, Suite 700

; CITY: San Diego

; STATE: California

; COUNTRY: USA

; ZIP: 92122

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/399,411

; FILING DATE: 06-MAR-1995

; CLASSIFICATION: 530

; ATTORNEY/AGENT INFORMATION:

; NAME: Campbell, Cathryn A.

; REGISTRATION NUMBER: 31,815

; REFERENCE/DOCKET NUMBER: P-LJ 1264

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (619) 535-9001

; TELEFAX: (619) 535-8949

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1719 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-399-411-4

Query Match 2.5%; Score 110.5; DB 2; Length 1719;

Best Local Similarity 19.1%; Pred. No. 0.46; Indels 333; Gaps 42;

Matches 169; Conservative 96; Mismatches 289; Indels 333; Gaps 42;

```

QY 78 FTSCSPENHFV-----IEIQKNDICMSGCPREGEVQLQSTSLPLTNTFTMDVKAHK 133
DB 358 FMFPCOCHCKERTTKQGLERHMH-----ISTVNAF-----KCK 393
QY 134 STGLELOFSIPRLQ-----IGPESCPDGV-----HSISGRIDATVVRIGT 176

```

Db 394 YCGAFGTQINRRRHERHHEAGLKRKPSQTLQPSEDLADKASGENVASKDDSSPSPISGP 453
QY 177 FC-----SNGTVSRIKMOEG-----VKMALHLPMFHRNVS GF 209
Db 454 DCLIMNSEKASQDTINSSVEENGEVKELEHPCCKYCKVFGTHMRRHQRVHERHLPK 513
QY 210 SIARRSSIKRLCLIESVEEGSATLMSANYPEGPEDE-----LMTWQFVYPAHLRA 262
Db 514 GVRKRGGLE-----EPQPPAQAOATQNVYVSTPEEBEGADVDYIMD----- 557
QY 263 SVSFLNLSNCKEERKEVEYIYG-----STNPEYFKLEDQPGMAGNFNLISLQGC 316
Db 558 -----ISSNIS-----ENLNYIDGKIOTNNNTSNCVIEHE----- 589
QY 317 DQDASPGILRLQFOVLVQHON-ESNKIYV-DLSNEBRAMSLTEPRPVKSRKFPVPGC 374
Db 590 SASADLYGINCLLPVVEITQNIKTQVPTEDLPKEPLGSTNSE--AKKRTASPA 646
QY 375 FVCLSESR-----CSSNLTLSGSKHKISFLCDDLRLMNVERTISTDHRICOR 426
Db 647 LPKIKAFETSDPMVPSCLSLPLSISTTEAVSF-----HKERSV-----YLSK 690
QY 427 SYSLOVPSDILHLPEVLEHDSMKLLVPKDRSLVLPQAQKLOQHTHEKPCTS SYLVAS 486
Db 691 LKOLLQTDKLTTPAGISATEIAKLP-----VCVSAPASML-----PVTSSRFKRTS 739
QY 487 AIPSDLYFGSFCGSGSIKQIVQKONISVTLRTFA-PSFOEA-SROGLTVSFTPYKEE 544
Db 740 SPSPSPQH-----SPALRDFGKPSDGAAMTDAGLT----- 770
QY 545 GVFTVTPDTKSKVYLRTPNMDRGLPSLTSVSNIS--VPRQVACLTFEKKRSQVVCOTG 602
Db 771 -----SKSKLSHSH-----DSPAMSLSGRDERETVSPCFDEKMKSEKWTAS 813
QY 603 RAFMIIOEQ-----RTRAEELFSLD--EDVLPKPSFHHHSFWNINCSPTSGKQ 650
Db 814 SAESVVCQDPLDSSGVKQKAEKGTKPVQWESYL-----DLSYHKHCHDSSEKKE 865
QY 651 LDLLFSVTLPRVTLVILIAAVGGVLLSALGLIICVKKKKKKTN--KSPAVGIY 707
Db 866 FKESHVSQPT-----CSAVKKRKPTTQMLQVLLNEY 897
QY 708 NG-NINTEMPR-----QPKKFKGKRDNDSHVAVIEDTVYGHLLQDSSGSLQPEVD 760
Db 898 NGIDLPEVNPADGTRSPSPCKSLAQDPD-----LQPGSG-PPAPVYE 940
QY 761 TYRPFQGTMGVCPSP-----PTICSR-----APTAKLATEEPP----- 794
Db 941 S-----TPDVCPSSPALQTPSLSSGQLPRLIPTDSSPPCPPVLTVAITPPPLPTV 994
QY 795 PRSPSESESEPTFESHNPNGDVSSKDTDIPLN-----TOBMEPAE 836
Db 995 PLPAPSSASBPHPCPSPISLN--ATAQSPILISPTVSPSPPIPPVE 1039

RESULT 5
; Sequence 4, Application us/08516859a
; Patent No. 6069231
; GENERAL INFORMATION:
; APPLICANT: Huang, Shi
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting
; TITLE OF INVENTION: Zinc Finger Proteins
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/516,859A
;; FILING DATE: 18-AUG-1995
;; CLASSIFICATION: 514
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/399,411
;; FILING DATE: 06-MAR-1995
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/292,683
;; FILING DATE: 18-AUG-1994
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Campbell, Cathryn A.
;; REGISTRATION NUMBER: 31,815
;; REFERENCE/DOCKET NUMBER: P-LJ 1776
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (619) 535-9001
;; TELEFAX: (619) 535-8949
;; INFORMATION FOR SEO ID NO: 4:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1719 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-516-859a-4

Query Match 2.5%; Score 110.5; DB 3; Length 1719;

Best Local Similarity 19.1%; Pred. No. 0.46; Mismatches 289; Indels 333; Gaps 42;

Matches 169; Conservative 96; Mismatches 289; Indels 333; Gaps 42;

QY 78 FTFCSCSPENHFV-----IELOKNIDCMGSPCFGEVOLQPSLPLTLNRTFINDVNAHK 133
Db 358 FMFCQCHERKFTTKQGLERMHNIH-----ISVYNNHF-----KCK 393
QY 134 STGLEQSLPRLQ-----IGPESCPDGYT--HSISGRIDATVVRIGT 176
Db 394 YCGAFGTQINRRRHERHHEAGLKRKPSQTLQPSEDLADKASGENVASKDDSSPSPISGP 453
QY 177 FC-----SNGTVSRIKMOEG-----VKMALHLPMFHRNVS GF 209
Db 454 DCLIMNSEKASQDTINSSVEENGEVKELEHPCCKYCKVFGTHMRRHQRVHERHLPK 513
QY 210 SIARRSSIKRLCLIESVEEGSATLMSANYPEGPEDE-----LMTWQFVYPAHLRA 262
Db 514 GVRKRGGLE-----EPQPPAQAOATQNVYVSTPEEBEGADVDYIMD----- 557
QY 263 SVSFLNLSNCKEERKEVEYIYG-----STNPEYFKLEDQPGMAGNFNLISLQGC 316
Db 558 -----ISSNIS-----ENLNYIDGKIOTNNNTSNCVIEHE----- 589
QY 317 DQDASPGILRLQFOVLVQHON-ESNKIYV-DLSNEBRAMSLTEPRPVKSRKFPVPGC 374
Db 590 SASADLYGINCLLPVVEITQNIKTQVPTEDLPKEPLGSTNSE--AKKRTASPA 646
QY 375 FVCLSESR-----CSSNLTLSGSKHKISFLCDDLRLMNVERTISTDHRICOR 426
Db 647 LPKIKAFETSDPMVPSCLSLPLSISTTEAVSF-----HKERSV-----YLSK 690
QY 427 SYSLOVPSDILHLPEVLEHDSMKLLVPKDRSLVLPQAQKLOQHTHEKPCTS SYLVAS 486
Db 691 LKOLLQTDKLTTPAGISATEIAKLP-----VCVSAPASML-----PVTSSRFKRTS 739
QY 487 AIPSDLYFGSFCGSGSIKQIVQKONISVTLRTFA-PSFOEA-SROGLTVSFTPYKEE 544
Db 740 SPSPSPQH-----SPALRDFGKPSDGAAMTDAGLT----- 770
QY 545 GVFTVTPDTKSKVYLRTPNMDRGLPSLTSVSNIS--VPRQVACLTFEKKRSQVVCOTG 602
Db 771 -----SKSKLSHSH-----DSPAMSLSGRDERETVSPCFDEKMKSEKWTAS 813
QY 603 RAFMIIOEQ-----RTRAEELFSLD--EDVLPKPSFHHHSFWNINCSPTSGKQ 650

Db 814 SASSVNCQNPRLDSSGVKAKAGTGTQVQMSVLT-----DLSVHKHCHSDSGKE 865
Qy 651 LDLESVTLTPRTVDTLVIIAAVGGVLLSALGLIICVKKKKKTN--KGPAGIY 707
Db 866 FKESHVQPT-----CSAVKKRKPTTCLQKVLINEX 897
Qy 708 NG-NINTEMPT-----QPKKFKGRKNDSHYAVIEDIMYVGHLLQDSSGFLOPEVD 760
Db 898 NGIDLPEVNADGTRSPSPCKSLAQPPD-----LGPSSG-FPAETVE 940
Qy 761 TYRFGQTMGVCPPSP-----PTICSR-----APTAKLATEEPP----- 794
Db 941 S-----TPDVCSSPALQTPSLSSGQLPPLILTPDSSPPPCPPVLTATPPPLLPV 994
Qy 795 PRSPSESEPTFFSHPNNGDVSSKDTDIPLN-----TQPMPEAE 836
Db 995 PLPAPSSASPHPCPSPLSN--ATAQSPILPILSTVSPSPSPPIPVPE 1039

RESULT 6

US-09-586-472-4
; Sequence 4, Application US/09586472
; Patent No. 633335
; GENERAL INFORMATION:
; APPLICANT: Huang, Shi
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting
; Zinc Finger Proteins
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/586,472
; FILING DATE: 01-Jun-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/528,706
; FILING DATE: 17-MAR-2000
; APPLICATION NUMBER: US 08/516,859
; FILING DATE: 18-AUG-1995
; APPLICATION NUMBER: US 08/399,411
; FILING DATE: 06-MAR-1995
; APPLICATION NUMBER: US 08/292,683
; FILING DATE: 18-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 4130
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1719 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-586-472-4

Query Match 2.5%; Score 110.5; DB 4; Length 1719;
Best local similarity 19.1%; Pred. No. 0.46;
Matches 169; Conservative 96; Mismatches 289; Indels 333; Gaps 42;

Qy 78 FTSCQSPENHFV-----IEIQKNIDCMGSPCPCEVQLQPSLTLLPTNRTFIWDVKAHK 133
Db 358 FMPCQCHCKERFTTKQGLERHMHIT-----ISTVNAF-----KCK 393
Qy 134 SIGLEQFSIPRLQ-----IGPESCPDGYT--HISGRIDATVVRIGT 176
Db 394 YCKAKFTQINRRHERHRAGLKRRPSQTLQPSBEDLADKASEENVASKDDSSPSLGP 453
Qy 177 FC-----SNGTVSRKKQEG-----YKALHLFWFHRNVSGF 209
Db 454 DCLIMNEKASOPTINSVVEENGEVKELHPCKYCKKVFQHTMRHRQVRHERHLIPK 513
Qy 210 SIARRSIRKLCITESTFEEGSGATILMSANTPEGFDE-----LMTQFVYPALRA 262
Db 514 GVRKRGLE-----EPQPAQAOATQVYVYVSTPEEEDADVYIMD----- 557
Qy 263 SVSFLNFNLSCERKEEREVEYIIPG-----STNPVYFKLEDQPGNMAGNFNLSTQGC 316
Db 558 -----ISSNIS-----ENLNTYIDGKIQTNNNTSNCVIEIE----- 589
Qy 317 DQDAQSPGILRLQFVLVQHPQN-ESNKIYV-DLSNERASLTIEPRPYKQSKRYVGC 374
Db 590 SASADLYGINCLLTPVTEITQNIKTQVPTEDLPKEPLGSTNSE--AKRRRTASPPA 646
Qy 375 FVCLSEST-----CSSNLTLGSGKKHISFLCDLTRLMMNVERTISCTDHRVYQK 426
Db 647 LPKIAKTDSDMPVSPCSLPLSISTEAVSF-----HKESV-----YLSK 690
Qy 427 SYSLOVPSDILHLPELCHDFESWKLVPKDRSLVLPVPAOKLQOHTHKPCNTSPSYLAS 486
Db 691 LKQLQIQODKLTPPAGISAEIAKLP-----VCYSASASML-----PTSSFKRTS 739
Qy 487 AIPSQDLVFGSFCGPGGSIKQIQVKNISVTLRFA-BSFOQA-SROGLTVSFLPYREE 544
Db 740 SPSPSPQH-----SPALRDEGKPSDGAATWDAGLT----- 770
Qy 545 GVFTVTDTKSKYVLRTPMNDRLGPLSLTSVMNLS--VPRQVACLFFKERSGVQCOTG 602
Db 771 -----SKRSKLESHS-----DSPAWSLSGDERETVSPCFDEKMKSEMTAS 813
Qy 603 RAFMIQEO-----RTFAEIFSLD--EDVLKPSPHHNSFWNINSCTSGKQ 650
Db 814 SASSVNCQNPRLDSSGVKAKAGTGTQVQMSVLT-----DLSVHKHCHSDSGKE 865
Qy 651 LDLESVTLTPRTVDTLVIIAAVGGVLLSALGLIICVKKKKKTN--KGPAGIY 707
Db 866 FKESHVQPT-----CSAVKKRKPTTCLQKVLINEX 897
Qy 708 NG-NINTEMPT-----QPKKFKGRKNDSHYAVIEDIMYVGHLLQDSSGFLOPEVD 760
Db 898 NGIDLPEVNADGTRSPSPCKSLAQPPD-----LGPSSG-FPAETVE 940
Qy 761 TYRFGQTMGVCPPSP-----PTICSR-----APTAKLATEEPP----- 794
Db 941 S-----TPDVCSSPALQTPSLSSGQLPPLILTPDSSPPPCPPVLTATPPPLLPV 994
Qy 795 PRSPSESEPTFFSHPNNGDVSSKDTDIPLN-----TQPMPEAE 836
Db 995 PLPAPSSASPHPCPSPLSN--ATAQSPILPILSTVSPSPSPPIPVPE 1039

RESULT 7

US-09-528-706-4
; Sequence 4, Application US/09528706
; Patent No. 6468985
; GENERAL INFORMATION:
; APPLICANT: Huang, Shi
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP

STREET: 4370 La Jolla Village Drive, Suite 700
City: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/528,706
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/516,859
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/292,683
FILING DATE: 18-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 1776
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1719 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-528-706-4

Query Match 2.5%; Score 110.5; DB 4; Length 1719;
Best Local Similarity 19.18; Pred. No. 0.46; Indels 333; Gaps 42;
Matches 169; Conservative 96; Mismatches 289;

78 FTFCSCSPENHFV---IEIOKNIDCMGSGPCEGVOLOPSTSLPTLMTFIMDVAKH 133
134 SIGELQSPRLRQ-----IGPESCPDGYT--HSISGRIDATVVRIGT 176
394 YCGKAFGTQINRRRHRHREAGLKRKPSQTLQPSDELADKASENVAASKDSSPSPISGP 453
177 FC-----SNGTVSRIKMOEG-----YKALHLPMFHRNYSGF 209
454 DCLIMSEKASQDTINSSVVEENGEVKEHLPCYCKVFGTHTMKRRHQRVHERHLIPK 513
210 SIANRSSIKRLCTIESVEEGSATLMSANYPEGFPDE-----LMTWQFVPAHLRA 262
514 GVRKRGLE-----EPQPAQAQAQATQNVYVSTPEEEGADVIMD----- 557
263 SVSLANLNLSCERKEVEYIYG-----STINPEYFKIEDKOPGMAGNENLISQGC 316
558 ---ISSNIS-----ENLNYIDGKIQTNNNTSNCVIEIEME----- 589
317 DQDQASGILRLQFOVLYVQHPN--ESNKIYV--DLSNERAMSLTIEPRPVYKSKRFYGC 374
590 SASADLYGINCILTPVVEITQNTKTOVPYTEDLPKEPLASTYSE---AKKRRTASDPA 646
375 FVCLSEST-----CSSNLTLTSGSKHKISFLCDDLTRLMNVERTISTCDHRYCQR 426
647 LPRKATFSDSPWVPSLSLPSISTTEAVSF-----HKESV-----YLSK 690
427 SYSIQVPSDILHLPEVLEHDSWKLVPKRLSLVLPQAKLOOHTHERPCNTSFSYLVAS 486
691 LKQLQTDQKLTLPAGISATEIAIKGP---VCVSAPASML-----PYTSSFFKRRTS 739
487 AIPQSDLYFSFCGSGSIKQIQOVKONISVTLRTFA--PSFOEA--SRQGLTYSFIPYKEE 544

740 SPSPSPQH-----SPALRDFGKPSDGAAMTDAGLT----- 770
545 GVFVTPDTKSKXYLRTPMWDRLGLPSLTYSWMNIS--VPDQVACLTFEKKERSGVYCQTG 602
771 -----SKKSKLESHS-----DSPAMSLSGRNERETVSPRCDEKMSKEMWAS 813
603 RAEMIOEQ-----RTBAEELFSLD--EDVLPKPSFHNNHSFWYINSCPTSGQ 650
814 SAFSSVNCQOPLDLSGKQAKEGTGKTPVQWESVL-----DLSVHKRHCSPSEKE 865
651 LDLFSTVTLPRVTDVLYILAAVGGVLLLSALGLIICVKKKKKTN--KGPANGV 707
866 FKESHVQPT-----CSAVKKRKPTCMQKYLNEY 897
708 NG-NINTEMPR-----QPKFKQGRKNDSHYAVIEDTWYHGLQDSSGSLQPEVD 760
898 NGIDLPEVNPADGTRSPCKSLAQRPD-----LQPSG-PPAPIVE 940
761 TYRPFQGTMGVCPSP-----PTCSR-----APTAKLATEEPP----- 794
941 S-----TPDVCPSPALQPTSLSSGOLPRLIPTDPSPPCPPVLTVAATPPPLPTV 994
795 PRSPSESEPTFSPHNPNGDVSSKDDIPLN-----TOBMPBAE 836
995 PLPAPSSASPHPCPSPLSN--ATAQSPPLTSPVSPSPPIPPVE 1039

RESULT 8
US-09-324-867-2
Sequence 2, Application US/09324867A
Patent No. 6251632
GENERAL INFORMATION:
APPLICANT: Lilliecrap, David
APPLICANT: Cameron, Cherie
APPLICANT: No. 6251632Jley, Colleen
APPLICANT: Horrocks, L. Suzanne Hoyle
APPLICANT: Hough, Christine
TITLE OR INVENTION: Canine Factor VII Gene, Protein and Methods of Use
FILE REFERENCE: 1669, 0010002/JAG/bjd
CURRENT APPLICATION NUMBER: US/09/324,867A
CURRENT FILING DATE: 1999-06-03
EARLIER APPLICATION NUMBER: 09/035,141
EARLIER FILING DATE: 1998-03-059
EARLIER APPLICATION NUMBER: 60/039,953
EARLIER FILING DATE: 1997-03-06
NUMBER OF SEQ ID NOS: 63
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 2343
TYPE: PRT
ORGANISM: Canis familiaris
US-09-324-867-2

Query Match 2.4%; Score 106.5; DB 4; Length 2343;
Best Local Similarity 18.28; Pred. No. 2; Indels 285; Gaps 38;
Matches 150; Conservative 108; Mismatches 283;

118 LPTL---NRTEIMDVAKHNSIGLEQSPRLRQIGPESCPD-----GYTHSISGRID 168
1030 IPTLLIENISVW-----QDIMEERNEFEKVTSLINHEFMNDMTALGLNH--VSNTK- 1082
169 ATVVRICTFGCSNGTVSRIKMOEGYKALHLR-----WFHPRANVSGFSIANRSS 216
1083 -----TLKNVMAHQKKEDPVLRAENPDLSKSKIPFLDWIKTHGNKSLSSQRP 1135
217 IKRLCII-----ESVFEGEGSATLMSANYPEGFPDELTMTQFVPAHLRA 262
1136 PKQTLTSGSESKYKQDQFLSEKVVGEDEFT--KTELQELFPPNNK----- 1180
263 SVSFLNF-----NLSCERKE-----ERVEYIIGSTTNP-----VFKLED 299
1181 SIFPANIANOENDTYOERKKSPEIERKEKLTQDENVALFOAHMIGTKNPLKMLFLIST 1240

QY 300 KOPNMG-----NPNLSIQ-----GCDQADSPGILRLQOVUWHPQ-----NESKTI--- 344
 Db 1241 KO--NWAGLEQPTPTLIQDTRSLNDSPHSGIIMANFSKIREANLEGLNQNOMWER 1238
 QY 345 -----YVVDLSNERAMSLTEPRPVKORKFVBCFVCELSRTCSNLTLYS 391
 Db 1299 FPRTRMSSNASHVITQGRKRSLK--QPRLSQGEIIF-----ERKVIANDSTQ 1346
 QY 392 GSKKHSFLCDLRLRMNNVEKITSIDHRY--CQRKSYSLQVPSDILHLEVELHDSWK 449
 Db 1347 WSKMNNYLAQGLTLQIEENKEKRAITQSPSLDCSMRN-----HVTIONDSA-- 1394
 QY 450 LLYPKDRLS-----LYLVPAOKLQDHTHEKPCPTSYLVASAIPSDQLPFSFCPG 502
 Db 1395 LPVAKESAPSVRYHTDILKTRISQNNSHLPASACNVTFRERTSGV-----OEG 1442
 QY 503 SIKOIQVKNISVTLKTRFAPSFOEASRQGLTVSEIPYFKEGEVETVY---PRTSKVY 558
 Db 1443 S-----HFLQEKARNNLSLAFVTLGITIEGQKRFESLSKSNATQPMY 1483
 QY 559 LRTPRMDRGLPSLTSVSNISVPRDOVACLTFPERKSGVVCQTGRAFMIOEDQTRAETI 618
 Db 1484 KLENTVLLQGLSETS-----DKVELLS-----QV 1509
 QY 619 FSLDEDVLPKPSFHNSFWWNISNCSPTSGKQDLLFSVTLPRTVDLYLLIAVGGGV 678
 Db 1510 HVDED-----SEPTYSNDSP--GHLDLM-----GKI 1535
 QY 679 LLLSALGLIICVYKKKKKKTN-----KGPA--VGI--YNGNINTEMPRO 718
 Db 1536 FLQKTQGV-----KMKRTKSPCKVPRFLKMAETSESEKIPSKLGLVLANMDHNYDTQITSE 1589
 QY 719 PKRFQKGRKNDSHVAYAVIEDTMV-YGHLQDSSGSLQPEVDYTRPQGTMGVCPSPSP 777
 Db 1590 EMKSQKKSQJTAFKR--KDTILPLGECENNDSTAAINEGOD--KPRGRAMAKQGEPRG 1644
 QY 778 TICGRAPYAKLATEPRPRSPSESEPTYSHPRNGVSSKDDI 823
 Db 1645 RLCSNPPVSKNHQREITVTTLQPEDEKFYDDTFSTEMKREDPDI 1690

RESULT 9
 US-08-473-750-11
 Sequence 11, Application US/08473750
 Patent No. 5834187
 Patent No. 5834187 5786143
 Patent No. 5834187 5786143
 GENERAL INFORMATION:
 APPLICANT: Green, Bruce A.
 APPLICANT: Brinton, JF., Charles C.
 TITLE OF INVENTION: Sequence and Analysis of LKP p11n
 Patent No. 5834187
 Patent No. 5834187 5786143
 TITLE OF INVENTION: Structural Gene and the LKP p11 Operon of No. 5834187 5786143
 TITLE OF INVENTION: Hemophilus Influenzae
 NUMBER OF SEQUENCES: 21
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
 STREET: Two Millitia Drive
 CITY: Lexington
 STATE: Massachusetts
 COUNTRY: US
 ZIP: 02173
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/473,750
 FILING DATE: 07-JUN-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/277,321

```

FILING DATE: 19-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Alice O.
REGISTRATION NUMBER: 33,542
REFERENCE/DOCKET NUMBER: ACC94-02B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 670 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-473-750-11

Query March                2.4%: Score 106; DB 2; Length 670;
Best Local Similarity      20.7%; Pred. No. 0.25; Indels 130; Gaps 18;
Matches 76; Conservative 49; Mismatches 113;

QY   53 AKPCYIVISKR-HITMLSIKSGERIVFTFSCSPENHFV-----IEIQ--KNIDCMGSP 103
    | | | | | : : : | : | : | : | : | : | : | : | : | : | : | : |
Db   413 ALPAIVASARDNDLTVHIOSEIAPVNSL-APENNIMTFKSAVELGYSFAGICTSNP 471

QY   104 CPFGVOL--QPSTSLPTLNRTFIWDYKAHKSTGLQLQSIPRLRIQPGESCDDGYTH 161
    || : : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db   472 CPTMKLPPLPHFLTLNLT------PGKKNSDG--- 497

QY   162 SISGIDATVRIRIFGCSNGTYSRIRIKMQGVKMAHLPMFHPRNVSGSIARRSIRKLC 221
    : : : : : : : : : | | | | | : : : | : | : | : | : | : | : | : |
Db   498 -----GEIRFLHNESING-----VSFOIGVKTNTSLDMVNKN-----NFSSLKVLM 539

QY   222 IIESVEEGSATLMSANYPEGPEDELMTWOVVPAHLRASVSL-NFNLSNCRKEER 280
    - - - - - | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db   540 V---PFNSKSISL-----HIRAKFHLLTDFSSLIAND----- 568

QY   281 VEYIYGSTTDEYFKLEDKOPGNMAGNFNISLGCCDDAOSPGILRLQFOVLVHPONE 340
    Y Y Y Y Y T P E Y F K L E D K O P G N M A G N F N I S L G C C D D A O S P G I L R L Q F O V L V H P O N E
Db   569 ----TTIDPMNTSICKIMLETWRGSTGNFSKYVGEDK----GDISIFF----- 609

QY   341 SNKIYVDLSNERAMSLT---IEPPVKQSKRFVGCCEVCLESRCSSMLNLTSGRKHKI 397
    : : : : : : : : : | : | : | : | : | : | : | : | : | : | : | : |
Db   610 -NTPKIIILKKOQRCTLNNAVPSPVX-----LRAVKKRLELAOSEMEGGTFLRV 659

QY   398 SFICDDLIT 405
    : : : : | : | : |
Db   660 N--CDNTT 665

RESULT 10
US-08-477-326-11
Sequence 11, Application US/08477326
Patent No. 5968769
GENERAL INFORMATION:
APPLICANT: Green, Bruce A.
APPLICANT: Britton, Jr., Charles C.
TITLE OF INVENTION: Sequence and Analysis of LKP Pilln
Patent No. 5968769
TITLE OF INVENTION: Structural Gene and the LKP Pill Operon of No. 5968769lypapb
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: US
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

```

RESULT 11
US-09-213-293D-1
: Sequence 1, Application US/09213293D
: Patent No. 6384299
: GENERAL INFORMATION:
: APPLICANT: GUTIERREZ-ARMENTA, CRISANTO
: APPLICANT: SANZ-BURGOS, ANDRES P.
: APPLICANT: XIE, QI
: APPLICANT: LOPEZ, PLATA S.
: TITLE OR INVENTION: PLANT RETINOBLASTOMA-ASSOCIATED PROTEINS
: FILE REFERENCE: 604-469
: CURRENT APPLICATION NUMBER: US/09/213,293D
: CURRENT FILING DATE: 1998-12-14
: PRIOR APPLICATION NUMBER: PCT/ES96/00130
: PRIOR FILING DATE: 1996-06-13
: PRIOR APPLICATION NUMBER: PCT/EP97/03070

RESULT 12
US-08-839-008-2
; Sequence 2, Application US/08839008
; Patent No. 5916758
; GENERAL INFORMATION:

APPLICANT: Hurler, Mark R
APPLICANT: McDonnell, Peter C
APPLICANT: McNulty, Dean E
APPLICANT: Rosen, Craig A
APPLICANT: Siemens, Ivo R
APPLICANT: Young, Peter R
TITLE OF INVENTION: Smooth Muscle Cell-Derived Migration Factor
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/839,008
FILING DATE: 23-APR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/563,697
FILING DATE: 28-NOV-1995
ATTORNEY/AGENT INFORMATION:
NAME: Baumeister, Kirk
REGISTRATION NUMBER: 33,833
REFERENCE/DOCKET NUMBER: P50384
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5096
TELEFAX: 610-270-5090
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 449 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-839-008-2

Query Match 2.4%; Score 104; DB 2; Length 449;
Best Local Similarity 20.6%; Pred. No. 0.19;
Matches 67; Conservative 41; Mismatches 102; Indels 116; Gaps 15;

151 PGESC-----PDGVTHSISGR-----DATV-----RIGTFC----- 178
59 PNKECAWTTTTPPEGQVYSLSFRVFDLELHPACRYDLAEVFAAGTSGQRLGRFCGTPRPA 118
179 -----SNGTVSRIMQEGVKMALHLPWFHPRNYSGFSIANRSSIKRLCITESVEEG---- 229
119 PLVAPENQVTLRMTDEGRGGREGFLMYSGRATSG-----TEHFGCGRL 164
230 EGSATLMSANYRPG-FPEDELMTQVFNPAHLRASVSFLNFNU--SNCERKEEREVEY- 284
165 KAGGTLTTPNWPESDYPGISCWMIAPPDQVIALTFEFKDELPPTYC--RYDSVSFVN 222
285 -----IFGSTNPVEFKLEDKQPGN-----MAGFNFLSL- 313
223 GAVSDSRRLGKFCGDAVPESTIS-----EGNELLVQFVSDLSVTADGFSASYKT 272
314 --QGCDQDASPGILRLQFVLVQHPONSNKIYVVDLSNERAMSLTIEPRPVKQSRKEV 371
273 LPRGTAKEGGPGPKRGTEPKVKVLPKSKQPE-----KTEESPASAPDAVTCRKQCR- 324
372 PGCFVCLBSRTGSSNLTLSGSKHKI 397
325 ---TGTLQSNFCASSLVMTATVSKMV 347

RESULT 13
US-08-475-844-9
Sequence 9, Application US/08475844
Patent No. 5972643
GENERAL INFORMATION:
APPLICANT: Lobanenko, Victor V.
APPLICANT: Neiman, Paul E.
APPLICANT: Klenova, Elena M.
APPLICANT: Filippova, Galina N.
APPLICANT: Collins, Steven J.
TITLE OF INVENTION: CTCF
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend Knourie and Crew
STREET: One Market Plaza
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/475,844
FILING DATE: 07-JUN-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/261,680
FILING DATE: 17-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Parmelee, Steven W.
REGISTRATION NUMBER: 31,590
REFERENCE/DOCKET NUMBER: 14538A-11-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-467-9600
TELEFAX: 415-543-5043
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 727 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-475-844-9

Query Match 2.4%; Score 103.5; DB 2; Length 727;
Best Local Similarity 16.4%; Pred. No. 0.51;
Matches 133; Conservative 99; Mismatches 261; Indels 319; Gaps 32;

166 RIAATVRICTGFSNGTVSR-----IKMOEGVKMALHLPWFHPRNYSGF 209
65 QLDPTLQMKTEVMEGVAPAEAAVDDQIITLQVNMEE-----QPINIGEL 113
210 SIAN-----RSSIKRLCITESVEEGSATLMSAN-----YPEGF----- 244
114 QLVQVVPVTVPVATTSVEEL--QGAIVENESKEGLAESEPMICHTLPLPEGFQVVKVG 170
245 -----PEDELMTQV--FVPAHLRASVSFLNPNNSNCERKEEREVEYIPG 267
171 ANGEVETLEOGLPPOQDPSPWQKDPDQPPAK-----KTKTKKSKLRTYEEG 218
288 STTNPEVFKLEDKOPGMMAGNFNLSLQGCDDQASPGILRLQFVLVQHPONSNKIYV 347
219 KVDVSVSYDEEBE-----QGBELLS----- 238
348 DLSNERAMSLTIEPRPVKQSRKRVPGCFVC-LBSRTC--SSNL-----TLTSGSKHKISF 399
239 EVNAEKRVGMMKPKPKTKIKKGVKKFQCELCSTCPRRSNLDRIHKKSHTDERRPK-CH 297
400 LCDDLTR--LMNVKERTICTDHRVQQRKSYSLQVPSDLHLPLVELHDSWKLVLVPKDR 456

```

Db 298 LCGRAFRVTLLRNHLNHTGTRPHKCPDCDMAFVTSGEIVR----- 339
QY 457 LSLVLPVAKLOQTHKEPCNTSFSYLVASAIPSDLYFGSCPGSIRKQIQVKNI--S 514
Db 340 -----HRRYKHTHEP-----FKSCMCYASVEVSKLRHITSH 373
QY 515 VLTLPFAPSGOQASROGLV-----SPIPY-----FKBEG--VFYTPDTKS 555
Db 374 TGEPRFOSLCSYASRDYTKLRMRHRTSGEKPYECYICHAFTQSGTMMKHILOKHTE 433
QY 556 KVIJRTPMWMDGLSLNSVSNISVPRQVA-----CLFFEKRSVVCOTGRAFI 608
Db 434 VAKHCPHCDTVIARKSDVLHKLKQHSYIEQKKCRYCDVAFHER-----YALI 483
QY 609 QEORT-RAEELFSLDE-----DVLPRKPSFHHSFWVINSICSPYSGKQ--LDLIFS 656
Db 484 QHQSINKNEKRFKDCDDYACROERHIMHKTHTGEPYACSHCDKFRKQOLDHMF 543
QY 657 VLTLPRTVDLTVILAAVGGVLLLSALGLIICVKKKKKTKNK-----GPAVG 705
Db 544 RYHDPNFPVPAFV-----CSKCGKFTFRNTMARHADNCAGP--- 580
QY 706 IYNGNINTEMROPKKQKGRK-----DNDSHYAVIEDTMYG 744
Db 581 --DGEENGEGETKRSKGRKRRKRSKEDSSSENAEPDLDDNEDEEPAVE----- 631
QY 745 HLLDSSGSFLQPEVDYRPFQGTMGVCPSPPTICSRAPAKLATEEPPRSPSESE 804
Db 632 -----IEPE-----PEPOVYTPRAPAKKRGRGPPORTNQPKNQ 666
QY 805 PYTSHPNNGDVSSKDDIPLNTQEPMEPAE 836
Db 667 PTAIOVEDONTGAIENTIVEVKKKEPDAEPAE 698

```

RESULT 14

PCT-0595-08429-9

Sequence 9, Application PC/TUS9508429

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: CTGF

NUMBER OF SEQUENCES: 21

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/08429

FILING DATE: 15-JUN-1995

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/261,680

FILING DATE: 17-JUN-1994

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Parmelee, Steven W.

REGISTRATION NUMBER: 31,990

REFERENCE/DOCKET NUMBER: 14538A-11-1PC

TELECOMMUNICATION INFORMATION:

TELEPHONE: 206-467-9600

TELEFAX: 415-543-5043

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 727 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

PCT-0595-08429-9

Query Match 2.4%; Score 103.5; DB 5; Length 727;
Best Local Similarity 16.4%; Pred. No. 0.51;

```

Matches 133; Conservative 99; Mismatches 261; Indels 319; Gaps 32;
QY 166 RIDATVNAIGTCFCSNGTVSR-----IMQGVNMAHLPMFHPNVSQF 209
Db 65 QLDPTLLQMKTEVMEGTAPAEAAVDDTQITILQVYVME-----QPINIEL 113
QY 210 SIAN-----RSSIKRLCITESVFEGSGATLSAN-----YBEGF 244
Db 114 QLVQVPVAVTVVATTSVEEL---QGAIVENVSKEGLAESPMICHTLPLDEGFQVAVG 170
QY 245 -----PEDELTWQ-----FVYPAHLASVSFLFNPINSCREBEREYILPG 287
Db 171 ANGEVETLEGGELPQEDPSNQKDDIOPAK-----KTKTKSKSLATYBEG 218
QY 288 STNPEVEKLEDKQPGNAGNFNLQGCDDAOSPGLRLFOYLQHPONESNKIYV 347
Db 219 KDVOVSVDPEE-----QOEGILS----- 238
QY 348 DLSNERASLTIERPVKQSKFVPGFCV-LESRTC--SSNL-----TLTSGKHKTSF 399
Db 239 EVNAEKVGNKPKPTKIKKGVKTFQCELCSTYCPRRSNLDRHMSKSHDERPHK-CH 297
QY 400 LCDLTLR---LMNVEXTISCTDHRQCRKSYSLQVPSDILHPLVELHDFSMKLLVPKDR 456
Db 298 LCGRAFRVTLLRNHLNHTGTRPHKCPDCDMAFVTSGEIVR----- 339
QY 457 LSLVLPVAKLOQTHKEPCNTSFSYLVASAIPSDLYFGSCPGSIRKQIQVKNI--S 514
Db 340 -----HRRYKHTHEP-----FKSCMCYASVEVSKLRHITSH 373
QY 515 VLTLPFAPSGOQASROGLV-----SPIPY-----FKBEG--VFYTPDTKS 555
Db 374 TGEPRFOSLCSYASRDYTKLRMRHRTSGEKPYECYICHAFTQSGTMMKHILOKHTE 433
QY 556 KVIJRTPMWMDGLSLNSVSNISVPRQVA-----CLFFEKRSVVCOTGRAFI 608
Db 434 VAKHCPHCDTVIARKSDVLHKLKQHSYIEQKKCRYCDVAFHER-----YALI 483
QY 609 QEORT-RAEELFSLDE-----DVLPRKPSFHHSFWVINSICSPYSGKQ--LDLIFS 656
Db 484 QHQSINKNEKRFKDCDDYACROERHIMHKTHTGEPYACSHCDKFRKQOLDHMF 543
QY 657 VLTLPRTVDLTVILAAVGGVLLLSALGLIICVKKKKKTKNK-----GPAVG 705
Db 544 RYHDPNFPVPAFV-----CSKCGKFTFRNTMARHADNCAGP--- 580
QY 706 IYNGNINTEMROPKKQKGRK-----DNDSHYAVIEDTMYG 744
Db 581 --DGEENGEGETKRSKGRKRRKRSKEDSSSENAEPDLDDNEDEEPAVE----- 631
QY 745 HLLDSSGSFLQPEVDYRPFQGTMGVCPSPPTICSRAPAKLATEEPPRSPSESE 804
Db 632 -----IEPE-----PEPOVYTPRAPAKKRGRGPPORTNQPKNQ 666
QY 805 PYTSHPNNGDVSSKDDIPLNTQEPMEPAE 836
Db 667 PTAIOVEDONTGAIENTIVEVKKKEPDAEPAE 698

```

RESULT 15

US-08-839-008-9

Sequence 9, Application US/08839008

Patent No. 5916758

GENERAL INFORMATION:

APPLICANT: Hurle, Mark R

APPLICANT: McDonnell, Peter C

APPLICANT: McNulty, Dean E

APPLICANT: Rosen, Craig A

APPLICANT: Siemens, Ivo R

APPLICANT: Young, Peter R

APPLICANT: Yue, Tian-Li

TITLE OF INVENTION: Smooth Muscle Cell-Derived Migration Factor

NUMBER OF SEQUENCES: 9

This Page Blank (uspio)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 11, 2002, 11:34:02 : Search time 36 Seconds
(without alignments)
334.830 Million cell updates/sec

Title: US-09-899-569A-4

Perfect score: 4394
Sequence: 1 MAGLNCGVSTALLGVLLGA.....SSKDRDIPLNQEMEPAE 836

Scoring table:

BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 92612 seqs, 14418503 residues

Total number of hits satisfying chosen parameters: 92612

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4394	100.0	836	US-09-899-569A-4	Sequence 4, Appli
2	3467	78.9	749	US-09-899-569A-2	Sequence 2, Appli
3	126	2.9	595	US-09-768-826-48	Sequence 48, Appli
4	114	2.6	477	US-09-972-715-9	Sequence 9, Appli
5	110.5	2.5	771	US-09-854-845-47	Sequence 47, Appli
6	110.5	2.5	870	US-09-854-845-31	Sequence 31, Appli
7	110	2.5	766	US-09-854-845-49	Sequence 49, Appli
8	110	2.5	865	US-09-854-845-33	Sequence 33, Appli
9	108.5	2.5	744	US-09-854-845-43	Sequence 43, Appli
10	108.5	2.5	843	US-09-854-845-27	Sequence 27, Appli
11	108	2.5	739	US-09-854-845-45	Sequence 45, Appli
12	108	2.5	838	US-09-854-845-29	Sequence 29, Appli
13	105.5	2.4	683	US-09-770-657-2	Sequence 2, Appli
14	105.5	2.4	683	US-10-023-676-1	Sequence 1, Appli
15	104.5	2.4	1042	US-09-888-615-74	Sequence 74, Appli
16	103.5	2.4	679	US-09-874-198-6	Sequence 6, Appli
17	103.5	2.4	679	US-09-874-238-6	Sequence 6, Appli
18	103	2.3	449	US-09-919-497-89	Sequence 89, Appli
19	103	2.3	458	US-09-925-301-1282	Sequence 1282, Ap

20	99	2.3	994	US-09-852-909-2	Sequence 2, Appli
21	98.5	2.2	1481	US-09-371-900-40	Sequence 40, Appli
22	98.5	2.2	1481	US-09-924-417-60	Sequence 60, Appli
23	97	2.2	556	US-09-888-615-106	Sequence 106, App
24	97	2.2	826	US-09-852-909-4	Sequence 4, Appli
25	97	2.2	2026	US-09-801-368-86	Sequence 86, Appli
26	96.5	2.2	937	US-09-949-192-19	Sequence 19, Appli
27	96.5	2.2	1884	US-09-785-770A-17	Sequence 17, Appli
28	96.5	2.2	1907	US-09-785-770A-16	Sequence 16, Appli
29	96	2.2	999	US-09-747-371-2	Sequence 2, Appli
30	95.5	2.2	3594	US-09-911-842-4	Sequence 4, Appli
31	95	2.2	906	US-09-746-491-48	Sequence 48, Appli
32	95	2.2	3034	US-09-737-149-25	Sequence 25, Appli
33	95	2.2	3034	US-09-737-149-30	Sequence 30, Appli
34	94.5	2.2	463	US-09-905-291A-285	Sequence 285, App
35	94.5	2.2	463	US-09-909-320-285	Sequence 285, App
36	94.5	2.2	463	US-09-909-088B-285	Sequence 285, App
37	94	2.1	535	US-09-768-826-38	Sequence 38, Appli
38	94	2.1	708	US-09-945-258-16	Sequence 16, Appli
39	94	2.1	811	US-09-992-598-57	Sequence 57, Appli
40	94	2.1	811	US-09-187-368-1	Sequence 1, Appli
41	94	2.1	811	US-09-989-722-57	Sequence 57, Appli
42	94	2.1	811	US-09-989-723-57	Sequence 57, Appli
43	94	2.1	811	US-09-989-279-57	Sequence 57, Appli
44	94	2.1	811	US-09-989-727-57	Sequence 57, Appli
45	94	2.1	811	US-09-989-731-57	Sequence 57, Appli

ALIGNMENTS

RESULT 1

US-09-899-569A-4

Sequence 4, Application US/09899569A

Patent No. US20020142003A1

GENERAL INFORMATION:

APPLICANT: No. US20020142003Albert Schweitzer

APPLICANT: Marva Scherl-Mostageer

APPLICANT: Wolfgang Sommergruber

APPLICANT: Roger Abseher

TITLE OF INVENTION: Tumorsozialer Antigen (B345)

FILE REFERENCE: 0652.2280001

CURRENT APPLICATION NUMBER: US/09/899,569A

CURRENT FILING DATE: 2001-07-06

PRIOR APPLICATION NUMBER: DE 100 33 080.0

PRIOR FILING DATE: 2000-07-07

PRIOR APPLICATION NUMBER: DE 101 19 294.0

PRIOR FILING DATE: 2001-04-19

PRIOR APPLICATION NUMBER: US 60/243,158

PRIOR FILING DATE: 2000-10-25

PRIOR APPLICATION NUMBER: US 60/297,747

PRIOR FILING DATE: 2001-06-14

NUMBER OF SEQ ID NOS: 40

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 4

LENGTH: 836

TYPE: PRT

ORGANISM: Homo sapiens

US-09-899-569A-4

Query Match 100.0%; Score 4394; DB 10; Length 836;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 836; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGLNCGVSTALLGVLLGAARLPKGAFAEIALPRSNITVILIKGTPTLLAKPCYIV 60

DB 1 MAGLNCGVSTALLGVLLGAARLPKGAFAEIALPRSNITVILIKGTPTLLAKPCYIV 60

QY 61 SKRHITMLSIKSGERIVFTSCSPENHFIETOKNIDKSGCPFGFVQLQSTSLPT 120

DB 61 SKRHITMLSIKSGERIVFTSCSPENHFIETOKNIDKSGCPFGFVQLQSTSLPT 120

QY 121 LNRFTMDVAKHSGISGLELQFSLPRLRQIGPGECSGPGVTHSISGRDIAIVRIGTFCSN 180

```

; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-899-569a-2
Query Match      78.9%; Score 3467; DB 10; Length 749;
Best Local Similarity 90.3%; Pred. No. 1,8e-292;
Matches 676; Conservative 9; Mismatches 28; Indels 36; Gaps 6;

QY 121 LNRFIFWDAHAKSIGLELFISIRLROIGEGSCPDGVTHSISGRIDATVVRIGTFCSN 180
QY 181 GTVASIKMOEGVKMALHPMFHPRNVSGETSIANRSSIKRLCIITISVEGEGSATLMSANY 240
Db 181 GTVASIKMOEGVKMALHPMFHPRNVSGETSIANRSSIKRLCIITISVEGEGSATLMSANY 240
QY 241 PEGPEDELMTWOFVPAHLRASVSFLNFMLSNCEKKEERYEYIIPGSTNPEVFKLEDK 300
Db 241 PEGPEDELMTWOFVPAHLRASVSFLNFMLSNCEKKEERYEYIIPGSTNPEVFKLEDK 300
QY 301 QPGMAENFMNLISQCCODQASPGILRLOFVLOHPONESNKIYYVDLSNERAMSLTIE 360
Db 301 QPGMAENFMNLISQCCODQASPGILRLOFVLOHPONESNKIYYVDLSNERAMSLTIE 360
QY 361 PRPVQKRKRVPGCFVCELESRTCSNLTLSGSKKISFLCDDLTFLMMNVEKTIISCTDH 420
Db 361 PRPVQKRKRVPGCFVCELESRTCSNLTLSGSKKISFLCDDLTFLMMNVEKTIISCTDH 420
QY 421 RYCQKRSYSLQVPSDILHLPVELHDFSMKLLVPKRLSLVLPQAOKLQOHTHEKPCNTSF 480
Db 421 RYCQKRSYSLQVPSDILHLPVELHDFSMKLLVPKRLSLVLPQAOKLQOHTHEKPCNTSF 480
QY 481 SYLVASAIPOSODLYFGSFCGSGSIKQIOVKONISVTLRTAPSFQOASROGLTVSEIPY 540
Db 481 SYLVASAIPOSODLYFGSFCGSGSIKQIOVKONISVTLRTAPSFQOASROGLTVSEIPY 540
QY 541 FKKEGVTVTPDFTKSKYYLRTPMNDGRLPSLTSVMNISVPRDOVACLTFEKKERSGVYVQ 600
Db 541 FKKEGVTVTPDFTKSKYYLRTPMNDGRLPSLTSVMNISVPRDOVACLTFEKKERSGVYVQ 600
QY 601 TGRAFMIIQORTRAEIEFSLDEVDLPKPSFHHSFWVNISNCSPTSGKOLDLFFSVTLT 660
Db 601 TGRAFMIIQORTRAEIEFSLDEVDLPKPSFHHSFWVNISNCSPTSGKOLDLFFSVTLT 660
QY 661 PRATVDLVIIAAGGVLLLSALGLITICVKKKKKTKNGAINGNINTEMPROPK 720
Db 661 PRATVDLVIIAAGGVLLLSALGLITICVKKKKKTKNGAINGNINTEMPROPK 720
QY 721 KFOGGRDNDSHYAVIEDTMVYGHLLQDSSGSFLQPEVDYRPFQGTMGVCPSPPTIC 780
Db 721 KFOGGRDNDSHYAVIEDTMVYGHLLQDSSGSFLQPEVDYRPFQGTMGVCPSPPTIC 780
QY 781 SRAPTAKLATEEPPRSPSESEPTTFSHPNNGDVSSKDTDIPLNTQEMEPAE 836
Db 781 SRAPTAKLATEEPPRSPSESEPTTFSHPNNGDVSSKDTDIPLNTQEMEPAE 836

RESULT 2
US-09-899-569a-2
; Sequence 2, Application US/09899569A
; Patent No. US20020142003A1
; GENERAL INFORMATION:
; APPLICANT: No. US20020142003Albert Schweitzer
; APPLICANT: Marva Scherl-Mostageer
; APPLICANT: Wolfgang Sommergruber
; APPLICANT: Roger Abseher
; TITLE OF INVENTION: Tumorsoziozietes Antigen (B345)
; FILE REFERENCE: 0652.2280001
; CURRENT APPLICATION NUMBER: US/09/899,569A
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: DE 100 33 080.0
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: DE 101 19 294.0
; PRIOR FILING DATE: 2001-04-19
; PRIOR APPLICATION NUMBER: US 60/243,158
; PRIOR FILING DATE: 2000-10-25
; PRIOR APPLICATION NUMBER: US 60/297,747
; PRIOR FILING DATE: 2001-06-14
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 749

```

```

; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-899-569a-2
Query Match      78.9%; Score 3467; DB 10; Length 749;
Best Local Similarity 90.3%; Pred. No. 1,8e-292;
Matches 676; Conservative 9; Mismatches 28; Indels 36; Gaps 6;

QY 67 MLSIKSGERIVTFPSCSPENHVEFIEIQKINDMSGCPGEOLOQSTSLPTLNRTFI 126
Db 1 MLSIKSGERIVTFPSCSPENHVEFIEIQKINDMSGCPGEOLOQSTSLPTLNRTFI 60
QY 127 WDVAHAKSIGLELOFSPIRLROIGEGSCPDGVTHSISGRIDATVVRIGTFCSN 186
Db 61 WDVAHAKSIGLELOFSPIRLROIGEGSCPDGVTHSISGRIDATVVRIGTFCSN 120
QY 187 KMOEGVKMALHPMFHPRNVSGETSIANRSSIKRLCIITISVEGEGSATLMSANY 246
Db 121 KMOEGVKMALHPMFHPRNVSGETSIANRSSIKRLCIITISVEGEGSATLMSANY 180
QY 247 DELMTWOFVPAHLRASVSFLNFMLSNCEKKEERYEYIIPGSTNPEVFKLEDKQPGMA 306
Db 181 DELMTWOFVPAHLRASVSFLNFMLSNCEKKEERYEYIIPGSTNPEVFKLEDKQPGMA 240
QY 307 GNFNLISQCCODQASPGILRLOFVLOHPONESNKIYYVDLSNERAMSLTIEPREVKQ 366
Db 241 GNFNLISQCCODQASPGILRLOFVLOHPONESNKIYYVDLSNERAMSLTIEPREVKQ 300
QY 367 SRKFPVPGCFVCELESRTCSNLTLSGSKKISFLCDDLTFLMMNVEKTIISCTDHRYCQK 426
Db 301 SRKFPVPGCFVCELESRTCSNLTLSGSKKISFLCDDLTFLMMNVEKTIISCTDHRYCQK 360
QY 427 SYSLOVPSDILHLPVELHDFSMKLLVPKRLSLVLPQAOKLQOHTHEKPCNTSFSLVYAS 486
Db 361 SYSLOVPSDILHLPVELHDFSMKLLVPKRLSLVLPQAOKLQOHTHEKPCNTSFSLVYAS 420
QY 487 AIPSQDLYFGSFCGSGSIKQIOVKONISVTLRTAPSFQOASROGLTVSEIPYFKKEGV 546
Db 421 AIPSQDLYFGSFCGSGSIKQIOVKONISVTLRTAPSFQOASROGLTVSEIPYFKKEGV 480
QY 547 FTVTPDFTKSKYYLRTPMNDGRLPSLTSVMNISVPRDOVACLTFEKKERSGVYVQGTGAFM 606
Db 481 FTVTPDFTKSKYYLRTPMNDGRLPSLTSVMNISVPRDOVACLTFEKKERSGVYVQGTGAFM 540
QY 607 IIOQORTRAEIEFSLDEVDLPKPSFHHSFWVNISNCSPTSGKOLDLFFSVTLPRATVDL 666
Db 541 IIOQORTRAEIEFSLDEVDLPKPSFHHSFWVNISNCSPTSGKOLDLFFSVTLPRATVDL 600
QY 667 TVIIAAGGVLLLSALGLITICVKKKKKTKNGAINGNINTEMPROPKPKOKR 726
Db 601 TVIIAAGGVLLLSALGLITICVKKKKKTKNGAINGNINTEMPROPKPKOKR 659
QY 727 KDNDSHYAVIEDTMVYGHLLQDSSGSFLQPEVDYRPFQGTMGVCPSPPTIC 786
Db 660 -----ERTMT--PWCQOSSRT-----PW- MGICYRIPAPSCQRT 694
QY 777 PTICSRAPAKLATEEPPRSPSESEPTTFSHPNNGDVSSKDTDIPLNTQEMEPAE 836
Db 695 PTICSRAPAKLATEEPPRSPSESEPTTFSHPNNGDVSSKDTDIPLNTQEMEPAE 723

RESULT 3
US-09-768-826-48
; Sequence 48, Application US/09768826
; Patent No. US20020012966A1
; GENERAL INFORMATION:
; APPLICANT: Shi et al.
; TITLE OF INVENTION: 18 human secreted proteins
; FILE REFERENCE: PF512P1
; CURRENT APPLICATION NUMBER: US/09/768,826
; CURRENT FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: PCT/US00/22350
; PRIOR FILING DATE: 2000-08-15

```

PUBLICATION NUMBER: 60/148,152
 PUBLICATION DATE: 1960-04-01
 PUBLISHED BY: HOMO sapiens
 PUBLISHED IN: HOMO sapiens

Query Match 19.98; Pred. No. Matches 24; 1100
 Similarity 2.98; n, mismatches 1100
 Best Local 135; Conservative 1100
 NCANVDCGPEDELMTWQFVPAH--LRAVSFLNFTNSNCERKERV 201
 -----DHV 70

```

OY      225 GVVGEQSSATLMSANRITCO-----GWTASDPVHGHWENRAG-----341
        :| | : | | :| | :
Db      33  TWEGLCVSYLCSEFSYPON-----KRLQFOVLVOHPONES 341
        :| | : | | :| | :

```

0Y 262 EYIPGSTINPEVFKLEDNF-----TRDRPH 2--
|| : || :
71 SRNIPATINNPAAVOEE-----CLESRRCSNLTTTSGSKHISFL 40
| : : |

0Y 342 NKIYVVDLSNERAMSLTIEPRVKOSN : : : : :
: SIRDRESDACTVEFCVERGNKKMKNYKXDLSTV
10A CTL-----LPV-----ELHDF 44

103 : : : :
DB CDDTLRLMNNVEKITSCTDHRRCAN : : : :
401 : : : :
QY : : : :
145 ODLSRYRLKPEVSIV-----QEGCLVSVPCSVLYPHYNWTASSPVYCSWFEKGADV : : : :
-----EKPCNTSFSLVVASAIPBODLYGCF 4 : : : :
2

[illegible][illegible]

```
Db      250 VEROSKNNMLT----- : ||| |
        VFTVTPD-TKSXYLRTPN-WDRGLPSLT----- : ||| |
OY      546 -----VTWTPD-TKSXYLRTPN-WDRGLPSLT----- : ||| |
        |::||| : | | | |
        TSCQVTLPGAVTMTRAVKLINISYPON-----
```

[illegible][illegible][illegible][illegible]

523 -----DANAVRG---SAGV-
 793 PPPRSPPESEPTESHPN 812
 //

573 ARPOLPOQEAIGYEYSEIN 594

RESULT 4
US-09-972-715-9
Application US/09972715
; sequence 9, Application
US20020110862A1

```
; Patent No.: 08796  
; GENERAL INFORMATION:  
; APPLICANT: Fousias, George  
; Inventor: Diamandis, Eleftherios  
; TITLE INVENTION: NOVEL SIGLEC-LIKE GENE
```

Page 3

PRIOR FILING DATE: 2000-10-06
PRIOR FILING DATE: 2000-10-06
NUMBER OF SEQ ID NOS: 15
SOFTWARE: patentin version 3.1

ORGANISM: *Neotoma*
S-09-972-715-9

2.68: Score 114; DB 10; Length 477;

[illegible][illegible]

81 MDIPVAIN-----QASRQGVSTPIPYKEKG-----
500 PGSLK-OIOVKONISVTLRTPAFSFO- : : : : :
||| : : ||| : :
...CUNATATHMPTFSLPGTLSECHPRNLCS-VPMACEGTPTTW 1

```
Db      133 ERGSRKWNYYIDKLVSATVIR...SVSNWISVPRDGV 5
OY     ---VFETPD--TKSKVYLRTPN--WDGRGLPSLT-----:||||:
          |::| |:| : --PDDHNG-TSLTQCQVLLPGAGVTMTAVRLNISTYPONTL 2
```

[illegible]

251 TMVFGGGDGTASTTLNGSALSLVGGSLHL--VGN--
625 ---VLKPSFH--HSEFWVNISNCSPTSGKQDLDFSV--TLTPRTVDLVILIAAVCG

309 NLGVLELPRVHVKDEGETCRAQN--PLGSHQHSLSLSLN-----
Db
677 G---VLLISALGLLICCVKKKKKTKNGKPAVGIGNNINTEMROPKKKQKGRKNDSDH

077 : | : : : :
| : : : :
367 ACATALLFLYFCITFVVVRSCRRKSAR-PAGCV--GDTGME---
Db
201 VATEDTMYVGHLLDSSGSFLQPEVDITYRPFQSTMGVCPSPPTICSRAPTKLATEEP

```
QY      /34 IAVLUD:--- | | :--GEIVINDS-  
        :    |   | :--  
Db      405 ----DANAVRG--SASOGPLIESPADDSPPHHAPALATPSPEE  
          -----DANAESEBYTESHPN 812
```

OY	794	P P R S P E S E A L T	:	:	:	:	:
		:	:	:	:	:	
Db	456	R P O Y P O E Q A I C Y E Y S E I N					474

RESULT 5
US-09-854-845-47
Sequence 47, Application US/09854845

```
; sequence: US20020098493.txt  
; Patent No.:  
; GENERAL INFORMATION:  
; Inventor: WALKER, D. Wade  
; APPLICANT: Wang, Xiaoming  
; Applicant:
```

APPLICANT: SCOVILLE, John
APPLICANT: Alexander Jr.
APPLICANT: Turner, C. 9520020098491a1el Human Semaphorin Homolo
TITLE OF INVENTION: No. US20020098491a1el
REFERENCE: LEX-0177-USA

FILE REFERENCE NUMBER: US/03/05-18
CURRENT APPLICATION NUMBER: 2001-05-14
CURRENT FILING DATE: US 60/205,274
CURRENT APPLICATION NUMBER: US 60/205,274
PRIOR APPLICATION NUMBER: 2000-05-18
Date: 2000-05-18

PRIOR FILING DATE: 06/20/2008
PRIOR APPLICATION NUMBER: US 60/208,933
PRIOR FILING DATE: 2000-06-02
PRIOR FILING DATE: 2000-06-02
NUMBER OF SEQ ID NOS: 50
Number of windows Version 4.0

```

NUMBER: 101
SOFTWARE: FASTSEQ
SEQ ID NO 47
LENGTH: 771

```

Matches 108, Conc

Query Mat
Best Loca
Matches

Query Mat
Best Loca
Matches

Query Match
Best Local
Matches 108; Similarity 2.5%; Score 110.5; DB 10; Length 870;
35; 403 DLTLMNVEKTI 15.88; Pred. No. 0.47; Mismatches 183; Indels 171; Gaps

DB 357 DQSRKRGTEGGVPEPPRPCTDLSRSGYNSODDLSTVLDE-VKHLPLARPVPT 455

QY 456 RLSLVLPADKLOQHHEKPCWTSFSTYLVASAFISODLYGSCFPGSGIKOYQWONISV 515

DB 416 GRELL- - - - -KRIRYTH- - - - -LTGFTVTPAGFTYDILLGTDGDMHNAVVLGSGMII 467

QY 516 TLTTFAPSRFOEASRGLGVSEFPF- - - - -KEEVEFTVTPTDPTK- - - - -VYLRTP- - - - - 562

DB 468 IEET- - - - -QVRESQVENVISLLOHSLYGCARSGYQLRLPLSCSRYSRCDICILARDPIC 525

QY 563 MNDRLGLSTLSVSNISVPRADYACLTFFEKESGVCOOTGAEMIIQEDONTAEETFSID 525

DB 526 GMPRGTHACAAATTIANNSQSRITALIODEKRGCESSRDTGPPPLKTRS- - - - -VLRGD 583

QY 623 EYVL- - - - -KPSFHHSFV- - - - -VNISNCSPP-SG- - - - - 648

DB 584 DVLRLCDQPSNLARALMILNLSKGLSDQGGTGVGDGL- - - - - 648

QY 649 -KOLDLTSTVLTPRYD- - - - -VNISNCSPP-SG- - - - - 648

DB 644 LRTLLASTSLTPRPATPAPAPAPATPGAO LAPDVRLLYLAIALALGGLCILIASCLLY 703

QY 689 CQYKKKK- - - - -LTVILIAVGGVLLSLALGII 688

DB 704 AGLNEGRRGRKRYSLGRASRAGSGASVQLQVTSGRALQVHNGMSPPAMPVLDGEPET 763

QY 714 EMRPQKKRKQKGRKDNDSEHYAVETEDMYG- - - - -KTNKGRPVGVYGNIN- - - - - 703

DB 764 QVLCQPR- - - - -PCVISHAHMECLASGLQCGH- - - - -HLDQSGSTLOPEV- - - - - 763

QY 768 IMCYQSPPTCTCSNAP- - - - - 763

```

0Y      :GVSQSLPHPIWSSSPAPCGDLFVKSLOTGPGEVRLHSPPLPSCVALYNOPPS---863
      DB    805 PYTES 809          :TAKLATEED-----CFIPASC 807
      864 PMSFS 868          :--P-----RSPSESEE 804
RESULT 7
US-09-854-845-49
; Sequence 49, Application US/09854845
GENERAL No. DE20020098491A1
APPLICANT INFORMATION:
APPLICANT: Waikie, D. Wade

```

APPLICANT : Scoville, Alvaoming
TITLE OF INVENTION : Turner, C. Alexander Jr.
FILE REFERENCE : LEX-0177-USA
CURRENT APPLICATION NUMBER : US-020020098491A1el
PRIORITY FILING DATE : 2006
PRIOR APPLICATIONS : Human Semaphorin

[illegible]

[illegible]

Query Match	2.5%	Score 108	DB 10	Length 739
Best Local Similarity	18.7%	Pred No. 0.59	Indels 188	Gaps 22
Matches	94	Conservative 73	Mismatches 149	
QY	403	DLTRLMNVNERTI-----SC-TDHRVCCORAKSYLQVPSDILHLPELHDFGSKLTVLPKD	455	
Db	258	DSSRRKRGVEGVPEPRPSCITDLSRSGVSSODPLVLADF-VKHLPLMARVPVPR	316	
QY	456	RSLVLYVPAOKLQOHTHEPCMTFSYLASAIPSDQLFSGFCFSGSKIOVKONIVY	515	
Db	317	GRPLL--KRNIKRYTH-----LTGTPVTPPAGPYRDLFLTAGDMHKAVVLCSGMI	368	
QY	516	TRTRAPSPFOCASRQGLTVSEIPF---KAGEVTVLPDRKSK-----VYLRTP--	562	
Db	369	IEET--OVERESQVENLYLSLQHSILVAGNAPSVGIOLPLSCSRYSRCICILANDPYC	426	
QY	563	NMDRG---LPSTLSVSNINISVPRDOVACLLTFPKKSGVVCOTGRAPMIIQOORTAELEF	619	
Db	427	GMDPETHACAAATTIANKTALIOD-----IERGNRCGESSRGTGPPPLKTRK--VL	476	
QY	620	SLDEDVLP--KPSFHHSFW-----VNISNCSTP-SG-----	648	

12-18:03:29 2002

DB 696 LVVACLRREGRR-----GRRKRYSLGRASR-----719
OY 746 LLDSSGSFLOPEVDTYRPFQCTMGVCP-----PSPPT 778
DB 720 ---AGGSANV-----LQTVSGQCPGEDEGDEGAGLGGSCLOIIPGEGAPAP- 766
OY 779 ICSRAPTAKLATEEPPRSPPESE 802
DB 767 -----PPPPPPPAE 776

DB 696 LVVACLRREGRR-----GRRKRYSLGRASR-----620
OY 649 ---KQDOLLRESVTLTPRTVD-----LTVLANAGSGVLLLSALG 685
OY 686 LIIICVKKKKKKTKNGKPAVGIYNGNINTEMPORPKKFGKRDNDSHYAVIEDTMYG 745
DB 397 LVVACLRREGRR-----GRRKRYSLGRASR-----620

DB 746 LLDSSGSFLOPEVDTYRPFQCTMGVCP-----PSPPT 778
DB 621 ---AGGSANV-----LQTVSGQCPGEDEGDEGAGLGGSCLOIIPGEGAPAP- 667
OY 779 ICSRAPTAKLATEEPPRSPPESE 802
DB 666 -----PPPPPPPAE 677

RESULT 12
US-09-854-845-29
Sequence 29, Application US/09854845
Patent No. US20020098491A1
GENERAL INFORMATION:
APPLICANT: Waik, D. Wade
APPLICANT: Alexander, J. C.
APPLICANT: Turner, C. Alexander Jr.
TITLE OF INVENTION: NO. US20020098491A1 Human semaphorin homologs and polynucleotids
FILE REFERENCE: LEX-0177-USA
CURRENT APPLICATION NUMBER: US/09/854,845
CURRENT FILING DATE: 2001-05-14
PRIOR APPLICATION NUMBER: US 60/205,274
PRIOR FILING DATE: 2000-05-18
PRIOR APPLICATION NUMBER: US 60/208,893
PRIOR FILING DATE: 2000-06-02
NUMBER OF SEQ ID NOS: 50
SOFTWARE: FASTSEQ for Windows version 4.0
SEQ ID NO 298
LENGTH: 238
TYPE: PRT
ORGANISM: homo sapiens
US-09-854-845-29

Query Match 2.5%; Score 108; DB 10; Length 838;
Best Local Similarity 18.7%; Pred. No. 0.73; Mismatches 149; Indels 188; Gaps 22;
Matches 94; Conservative 73;
DB 403 DUTRLMMVEKTI-----SC-TDHRVCQKSKSYSLQVSPDILHLPVELHDFSKLLVPRD 455
OY 337 DGRKMGVREGVPEPRPGSCITDLSRSGYNSODPLSLVDF-VKLPRMARPVVPR 415
DB 456 RLSLVLRQKLOQHTHEKPCNTSFSYLVASAIPSODLYGSGCPGSGIKOIVONKISY 515
OY 416 GRPLL-----KRWIRTH-----LTGFTVTPAGPTDILFLGRADGHIKAAVLGSGMI 467
DB 516 TLRTPAFSFOEASRQGLTFSEIYF-----KEEGFTVTPDYTKS-----VYLRTP-- 562
OY 468 IET--QVRESQSVENIVISLHSLYVGAPSGVIOLPLSSGSRYSQCDLIARDPYC 525
DB 563 MNDRG---LPSLTSVMNINVPDQVACITFFKERSGVVOTGAFAMIOEORTAREIF 619
OY 526 GMDPGTHACAAATTANTALIOD-----IERGNRGCSSTDTGPEPLKTRS--VL 575
DB 620 SLDEVLV--KPSFHHSHW-----VNISNCSPT-SG-----648
OY 576 RGDVLLPCDOPSLARALWILNGSLDGGGGRVNGVLDLYTDAQPEHSGVGYAE 635
DB 649 ---KQDOLLRESVTLTPRTVD-----LTVLANAGSGVLLLSALG 685
OY 636 ENGLRTLLASYSLVTRPATPAKAPATPGACQLAPDVRLLYVALAALGGLIASL 695
DB 686 LIIICVKKKKKKTKNGKPAVGIYNGNINTEMPORPKKFGKRDNDSHYAVIEDTMYG 745

DB 696 LVVACLRREGRR-----GRRKRYSLGRASR-----719
OY 746 LLDSSGSFLOPEVDTYRPFQCTMGVCP-----PSPPT 778
DB 720 ---AGGSANV-----LQTVSGQCPGEDEGDEGAGLGGSCLOIIPGEGAPAP- 766
OY 779 ICSRAPTAKLATEEPPRSPPESE 802
DB 767 -----PPPPPPPAE 776

DB 696 LVVACLRREGRR-----GRRKRYSLGRASR-----719
OY 746 LLDSSGSFLOPEVDTYRPFQCTMGVCP-----PSPPT 778
DB 720 ---AGGSANV-----LQTVSGQCPGEDEGDEGAGLGGSCLOIIPGEGAPAP- 766
OY 779 ICSRAPTAKLATEEPPRSPPESE 802
DB 767 -----PPPPPPPAE 776

RESULT 13
US-09-770-657-2
Sequence 2, Application US/09770657
Patent No. US20020046416A1
GENERAL INFORMATION:
APPLICANT: Gutierrez-Armenta, Crisanto
APPLICANT: Sanz-Burgos, Andres P.
APPLICANT: Xie, Qi
APPLICANT: Lopez, Paula S.
TITLE OF INVENTION: PLANT RETINOBLASTOMA-ASSOCIATED PROTEINS
FILE REFERENCE: guiteriez
CURRENT APPLICATION NUMBER: US/09/770,657
CURRENT FILING DATE: 2001-01-29
PRIOR APPLICATION NUMBER: PCT/ES96/00130
PRIOR FILING DATE: 1996-06-13
PRIOR APPLICATION NUMBER: PCT/EP97/03070
PRIOR FILING DATE: 1997-06-12
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 683
TYPE: PRT
ORGANISM: Unknown
FEATURE:
OTHER INFORMATION: Description of Unknown Organism: protein
US-09-770-657-2

Query Match 2.4%; Score 105.5; DB 10; Length 683;
Best Local Similarity 19.3%; Pred. No. 0.86;
Matches 146; Conservative 98; Mismatches 250; Indels 261; Gaps 38;
DB 109 VOLOPSTSLPTLN--RTFIWDVKAHKSIGLEQFSIPRLRQGPSPGDPGVTHSISGR 166
DB 88 VQMPVTSAMTAKMLREVISLDPKPSKIQ-QFL-----SSCRDLTNVTER 136
OY 167 IDATVVRIGTFCNGTYSRIKMGVKKALH-----LPMFHPRNVSGFSIANRSIKRL 220
DB 137 V--STVLEAIFPTSSANR-----GVSGLGNCANAFDIPMAEARKVASKLYR-VLEAI 188
OY 221 CITESVEFGESATIMSNYEGF-----PDELTMTQFVY--PAHLRASVSFLNF 269
DB 189 CRAELQNSNVNVLPLLN--ERFHCILACSADIVLATHKTVIMPPVYLE-STGLTAF 245
OY 270 NLSNCKEEREVEYIIPSTNPEYF-----LEDKOPGNAGFNLSLQCCDDQDASPG 324
DB 246 DLSKI-----LENVVRHETLPRKLKRLMSLEQLLESNAWEGSSLVN-----290
OY 325 ILRLOFVLVHPO--NENKRIYV-----DLSNRASLATEPPRVKSRKVPV- 373
DB 291 ---SLIVARSVASELNRGLLAEPNPSIDDLVSRG--NVREGLPAPFSKRAAGP 342
OY 374 -----CFVCLSERTC-----SSNLTLSGSKHKSIFLCCDITRIMN-- 410
DB 343 DNDADRSRKRSCNERNYVERNLQTPPKQSHVSTV-----LKAKCHPLQSTFASPT 397
OY 411 ---VEKTSICDTHKRCQKRSYSQVSPDILHLPVELHDFSKLLVPRDRLSLVPAOK 466
DB 398 VCPNPGVGNKCAP-----VTHHFFSKIL-----KLAAIR 427
OY 467 LQOHTHEKPC--WTSPSYLVASAIPSOD--LVFGSF-----CPGSGTKQIVQKONISY 515

Paragraph 1 Application US/10025676
 Patent No. US2002013847A1
 GENERAL INFORMATION:
 APPLICANT: GUTIERREZ-ARMENTA, CRISANTO
 APPLICANT: SANZ-BURGOS, ANDRES P.
 APPLICANT: XIE, QI
 APPLICANT: LOPEZ, PAULA S.
 TITLE OF INVENTION: PLANT RETINOBLASTOMA-ASSOCIATED PROTEINS
 CURRENT REFERENCE: 604-468
 CURRENT APPLICATION NUMBER: US/10/025 676
 PRIOR FILING DATE: 2001-12-26
 PRIOR APPLICATION NUMBER: 09/213, 294
 PRIOR FILING DATE: 1998-12-14
 PRIOR APPLICATION NUMBER: 09/213, 294
 PRIOR FILING DATE: 1996-06-11
 PRIOR APPLICATION NUMBER: PCT/EP97/03070
 FILING DATE: 1997-06-12
 OPTIMARE: Patenting No. 1997-06-12

[illegible]

290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533
 534
 535
 536
 537
 538
 539
 540
 541
 542
 543
 544
 545
 546
 547
 548
 549
 550
 551
 552
 553
 554
 555
 556
 557
 558
 559
 560
 561
 562
 563
 564
 565
 566
 567
 568
 569
 570
 571
 572
 573
 574
 575
 576
 577
 578
 579
 580
 581
 582
 583
 584
 585
 586
 587
 588
 589
 590
 591
 592
 593
 594
 595
 596
 597
 598
 599
 600
 601
 602
 603
 604
 605
 606
 607
 608
 609
 610
 611
 612
 613
 614
 615
 616
 617
 618
 619
 620
 621
 622
 623
 624
 625
 626
 627
 628
 629
 630
 631
 632
 633
 634
 635
 636
 637
 638
 639
 640
 641
 642
 643
 644
 645
 646
 647
 648
 649
 650
 651
 652
 653
 654
 655
 656
 657
 658
 659
 660
 661
 662
 663
 664
 665
 666
 667
 668
 669
 670
 671
 672
 673
 674
 675
 676
 677
 678
 679
 680
 681
 682
 683
 684
 685
 686
 687
 688
 689
 690
 691
 692
 693
 694
 695
 696
 697
 698
 699
 700
 701
 702
 703
 704
 705
 706
 707
 708
 709
 710
 711
 712
 713
 714
 715
 716
 717
 718
 719
 720
 721
 722
 723
 724
 725
 726
 727
 728
 729
 730
 731
 732
 733
 734
 735
 736
 737
 738
 739
 740
 741
 742
 743
 744
 745
 746
 747
 748
 749
 750
 751
 752
 753
 754
 755
 756
 757
 758
 759
 760
 761
 762
 763
 764
 765
 766
 767
 768
 769
 770
 771
 772
 773
 774
 775
 776
 777
 778
 779
 780
 781
 782
 783
 784
 785
 786
 787
 788
 789
 790
 791
 792
 793
 794
 795
 796
 797
 798
 799
 800
 801

: sequence 74
 : Patent No. US2002006485A1
 : GENERAL INFORMATION: 09/888615
 : APPLICANT: PLOWMAN, GREGORY
 : APPLICANT: WHYTE, DAVID
 : APPLICANT: CAENEPEEL, SEAN
 : APPLICANT: CHARYDCZAK, GLEN
 : APPLICANT: MANNING, GERARD
 : APPLICANT: SUDARSANAM, SUCHA
 : TITLE OF INVENTION: NOVEL PROTEASES
 : FILE REFERENCE: 038602/1214
 : CURRENT APPLICATION NUMBER: US/09/888,615
 : CURRENT FILING DATE: 2001-06-26
 : PRIOR APPLICATION NUMBER: 60/214,047
 : NUMBER OF SEQ ID NOS: 150
 : SOFTWARE: PatencIn Ver. 2.1
 : SEQ ID NO 74
 : LENGTH: 1042
 : TYPE: PRT
 : ORGANISM: Homo sapiens
 : 09-888-615-74

```

every Match
          2.43; Score 104.5; DB 10; Length 1042;
          152; Conservative 95; Mismatches 265; Indels 315; Gaps 39;
9  STALGVLLGAAKLPRGAAFEIALLPRESNITVLKLT-CNPILLAKPCIVYISKRAHTIM 67
   ::|:|::|:|  |  |||::|:|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
330 ALAVLSHMLSLQSDP---EAFHLIVPHVNLVHFSKNDGLPSSTA----- 372
68 LSIKSGRIIVTFSSQSPENHVIETQKNIDCM-----SGPCPFGEVQLD-----PSTS 116
   |:::|::|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
373 -----FLVQLTELIIHGMVHYSGFPDLYPELLEAIKIDPFKPSSE 411

```


This Page Blank (uspto)

Db 2585 PGYNGIRDYARNLDCEWTLSPNRENSSISYFLELSIESHQDCT----- 2622

```

Db 546 RYENLNDDNEPEPELPSALPIFQVFNPKNTSKPPIALIGLTLTRADDFSPITFIYHLLPNCGTPESSD 605
OY 324 GI-LRLOFQVILVHOQ-NESKKITIVDLSNBRASLTIEPRPVKQSKREVPGCCFVCLIESR 381
Db 606 NFENIDAEFGEIYVEPRKVDSPKSKTEAEV-----CFIATSOK 641
OY 382 TC-SSENLTITGSGSH-----KISPLCDLTLRLMMNVKETSCTDHRVCQQRKSYSLOQPSDI 436
Db 642 NLDTSEYFEDANSKNKFRKVEYFEGD-----SVDSVANSKGINNGFSGNOSITSSYGDV 684
OY 437 L--HLPEYL-----HDFSCKLLVPKDRLSLVLPQAQ-----KLOQTHIE 473
Db 695 LDRVEIPMNLGAAGADYEOK-----SLNFVPRANTYELGRDMISPEGAVELNKNTGE 746
OY 474 KPQNTSEYLVASA--IPSDLY---FGSFCRGSSTKOLOQVON-----ISVTLRTF 520
Db 747 -----LVANKILIDTPQGVYTAIEIGDGGEGEYKOLHNIRNRKRLRYVLSMNERNE 797
OY 521 APSEFOEASROGLTVSFITPYKEGVTTPYDPRS-----KVYLRTPNMDRGSLPSTVS 575
Db 798 GANLEK-----FRRIILAIKIDQOKAGKOLEIHFDEPKADRRKSTWTSV-- 842
OY 576 WNISVPRDQVACLTFPEKERSGVNCQGRAMTIOEOFTRAEETISLEDEVLPRSPFHHS 635
Db 843 -----CFYLLRONALIDENQASIISSPSNGHISKL-----HHI 875
OY 636 FMV-NISNCSP-----TSGKQDLDSVTLTPTVDTYILITAAVGGGVLLSLAG 685
Db 876 FKVDWDMCAKRPATISESTSSNSDIPLN-----TLLIGVGLLITALLAL 925
OY 686 LIICVAK-----KKKTKN--GPAVGIYNGNINTEMPROP 719

```

RESULT 3
A:30788
mannose 6-phosphate receptor protein, cation-independent - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A25908; A30788; S09404
R:lobel, P.; Dahms, N.M.; Kornfeld, S.
J. Biol. Chem. 263, 2563-2570, 1988
A:Title: Cloning and sequence analysis of the cation-independent mannose 6-phosphate
A:Reference number: A92706; MUID:88115411; PMID:2965004
A:Accession: A25908
A:Molecule type: mRNA
A:Residues: 1-2499 <LOB>
A:Cross-references: GB:J03527; NID:g162873; PIDN:AAA30455.1; PID:g162874; GB:M15869
R:Glickman, J.N.; Conlhard, E.; Pearce, B.M.F.
EMBO J. 8, 1041-1047, 1989
A:Title: Specificity of binding of clathrin adaptors to signals on the mannose-6-phos-
A:Reference number: S09404; MUID:89305502; PMID:2545438
A:Contents: annotation; HA-II adaptor binding

C; Keyword

```

F:1252-1532/Domain: hivonecncr type II repeat homology <ref>
F:12360-2365/Region: HA-II adaptor binding

Query Match      2.7%; Score 118; DB 1; Length 2499;
Best Local Similarity 19.9%; Pred. No. 11;
Matches 195; Conservative 100; Mismatches 343; Indels 340; Gaps 56;

OY      1 MAGLNCVSIALLGVLLIGAAKLRPRGAEEFIALPRESENTIVLKLGPPTLL---AKPC 56
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      1710 MHGIACPGACTAVCKPVDPG-----PIDIGRVA GPILNPDIANEVYLNESSTPC 1759

OY      57 YIVISKRIITMAISKSGRIEYFTESCO-----SPE-----NHVEIOLKNIIDC---M 100
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      1760 ---LADRIENTSL-----ITFHCKRGVSMGPKPLKLRISYCDVEFMEEHPILVCPDEVK 1809

OY      101 SGCPPEFGEVQLQPSSTLLPTLNRTFIWDVKAHK-SIGLEQFSIPRLQIGPGE-SCPDG 158

```

```
Db 1810 TDGCSLTDEQLYYSPNLSLSKSTFKVTRGPHYISGV-----CTAAAGLDEGGCKDG 1862
Qy 159 VTHISGRDIAIVVIGTFCNSGTVSRIMQEGVKMALHLPFHRPNVSGFSTANRSIK 218
Db 1863 AVCLLSGKSGAFGRLA-----SMKLDLRHDEAVILSYLANDTDP 1903
Qy 219 -----RLCTIISVFESEGSATLMSANYPEGPEDELMWQVFAHNLASVSLFNLS 272
Db 1904 PETEDGEPCVFPFVNGK-----STEECVESARALW-----CATTANYDD 1945
Qy 273 N-----CERKEERYEYIIPGSTNPEVFKLEDKOPGNMAGNFMLSLOGCDQDASPILRL 328
Db 1946 HEMGCKHSTSH-----RTSVIIFK-----CEDAD----- 1971
Qy 329 QPQVLYQHONSKKITYVVDLSNERAMSLTIERR-----PKQSKRY----- 371
Db 1972 -----VGRPO-----VFSEVRGCEVTPEMKTKVCPKRMCKEYQKHRTYDLRL 2017
Qy 372 ---PGCF-----VCLF-----SRTGSSNLT-----TSGSKHKISF-----LC 401
Db 2018 SLLTSMWSFVHNGASTYINLCKITKGPDCSERASVCKSTSGEVQVGLVHTOKLDV 2077
Qy 402 DDLRLMNVETICTDHRVCO-----RKSYS-LQVPSDILHPELHDFSM 448
Db 2078 DD-RVITYYSGKHGCGDKKTSAVIELCAKTVGRSPFTREDVDSCTYH-----FSM 2128
Qy 449 KLLVPRDLSTLVPAQKIQ--QHTEKPCN--TSESYLVAISPQDLYFGSCPGSGIK 505
Db 2129 -----DSRAACAVRPOEVQVNGTITNPANGRSFSL-----GDIYFKRFSASGDVR 2174
Qy 506 -----QIOYK-----QNISVTLRTFAPSFQOASRQGLTVEFIYFKEEG-- 545
Db 2175 TNGDRYIEIQLSTGSSSPACSGASICQ--KANDHFGRKVTGTSQTRYYQDGD 2231
Qy 546 --VFT-----VTPDTKSKYLLR-TPMMDRGLPSLTGSVNINISVPROVACLFFPKR 594
Db 2232 DVFYSSSSKCGDKKTSVSTIEFHODPLVKDGIPEFS-----HETACQYLFSMH 2282
Qy 595 SCVVCOTGAEMTIOBQRTABEILSLDEDVLPKPSFHHHSFVNINISNCSPTSGKQDIL 654
Db 2283 TSAVPLGAGF-----DEIAGDDAQEHK-----GLSERSQAVGAILSL 2322
Qy 655 FSVTLPRVVDLTVILIAVGGVLLLSALGLIC-----VKKKKKTKNKPRAVITNG 709
Db 2323 ---LVALTACLLTLTLTKERREVMRSR--LTNCCRSANSVKSKVKNKEEA---D 2372
Qy 710 NINTE-----MPROPKFKQGRKDNDSHYAV--IEDTM--VYGHLLQDSSGFLQ-P 757
Db 2373 EMETEMLMELIOPRAPRPGKEGE--NGHVAAKSVAAADTLGALHD--EQDSEDEVLTLP 2429
Qy 758 EVDYTRPFO--GTMGVCPSPPTTCSRAPLAK-----LATEPPRSPSESESEPTYS 809
Db 2430 EKVVRPPGAPGAEQ--GPRLRPLPRKAPPLRADRDVGLVRGPAPRGRAPRAAATPISF 2488
Qy 810 HPNNDGVSSKQDIDPLN 827
Db 2489 H-----DSDDEDLH 2498
```

RESULT 4

T09456
Intrinsic factor-B12 receptor Cubillin precursor - human
C:Species: Homo sapiens (man)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 02-Aug-2002
C:Accession: T09456
R:Kozuyaki, R.; Kristiansen, M.; Sllabharoglu, A.; Hansen, C.; Jacobsen, C.; Tommerup, N
Blood 91, 3593-3600, 1998
A:Title: The human intrinsic factor-vitamin B12 receptor, cubillin: Molecular characteriz
ion.
A:Reference number: 216777; MUID:98241400; PMID:9572993
A:Accession: T09456
A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA
A:Residues: 1-3623 <K02>
A:Cross-references: EMBL:AF034611; NID:93929528; PIDN:AAC82612.1; PID:93929529
C:Genetics:
A:Map position: 10p12
C:Superfamily: intrinsic factor-B12 receptor cubillin; EGF homology
C:Keywords: receptor; vitamin B12 uptake
F:1-24/Domain: signal sequence #status predicted <Sig>
F:25-3623/Product: intrinsic factor-B12 receptor #status predicted <Mat>
F:436-467/Domain: EGF homology <EGF>

Query Match 2.7%; Score 118; DB 2; Length 3623;
Best Local Similarity 19.2%; Pred. No. 20;
Matches 160; Conservative 84; Mismatches 271; Indels 320; Gaps 37;

```
Qy 88 HVEYIEIQNKIDC-----MSG---PCPGEVOLQPSLTSLPTIKTF--TWYKAHKSIG 136
Db 2352 HPTLPYRDMLFCMHLQSGHYLTLSFEDFNQNSG-----CEKDFVEIMDNHTSGNI- 2406
Qy 137 LELQFIPRLRQIGPESCPDGYTHSISGRIDATVYRIGTFCNSGTVSRIMQEGVKMAL 196
Db 2407 -----LGRYCGNTIPDSIDTSSNTAVR--FVIDGSVT----- 2437
Qy 197 HLPWFRPNVSGFSIANRSIKRLCTIISVFESEGS-ATLMSANYPEGPEDELMWQV 255
Db 2438 -----ASGRRLFESSME-----ECGGDQSGSIGFTSPNYPNPNPHGICMRIT 2483
Qy 256 VPAHNLASVSLFNLS---NCEKKEERYEYIIPGSTNPEVKL-----EDKOPGN 304
Db 2484 APEGRITLTMFNRLATHPSC--NNEHYLVFNGIRSNSPQLEKLSVAVNSEIKSSG 2541
Qy 305 MA-----GNFNLSLOGCDQDQSPGLRLQFQVLYQHONSKKITYVVDLSNE 352
Db 2542 TMKVIFFTDGSRPYGGFTASYS--SEDAVCGGLPMTPEKNFTSPYDGVANRNLNCE 2600
Qy 353 RAMSLTIERRPVYKSKKFPVGCLESRTSSNLTLTSGSKKISFLCDDTLRLMANYE 412
Db 2601 WTLG-----NNOGNSISIHFEFYLE----- 2623
Qy 413 KTISCTDHRVCOKRSYSLQVPSDILHPELHD--FSMKLLVPRDLSTLVPAQKIQH 470
Db 2624 -----SHQDCQ-----FDVLEFRVGDADGAPLMWRGSPKPLPLPIYPYSGWTH 2668
Qy 471 --THEKPCMTSF--SYLVASATPSQDLYFGSPGGSISQIQVKQINISVTLRTFAASFQ 526
Db 2669 FVTNERVEHIGFAKY-----SFTDCGGIQ----- 2693
Qy 527 EASRQGLTVSFIPYFKEEGVFTVTPDTKSKYLLRTPMMDRGLPSLTGSVNINISVPRDQVA 586
Db 2694 -----IGDSGYIT-----SENYPRAYDSLTHCSMLEAPRGHTI 2727
Qy 587 CLTF--FKERSGVVC-----VTLPRTVVDLTVILIAVGGVLLLSALGLICVKKKKK- 608
Db 2728 TLTFSDFLPEPHTTCAWDSTVYRANGSPESPITIGQCGNSNPRTIQSGSNQVLVTNSDH 2787
Qy 609 -----QOQRTABEILSLDEDVLPKPSFHHHSFVNINISNCSPTS---GKQDL 653
Db 2788 SLQGGFYATWNTQTLCGGCIFHSDNGTIRSP--HMPONFENPNSRCSWATITKSHLEI 2845
Qy 654 LFS-----VTLPRTVVDLTVILIAVGGVLLLSALGLICVKKKKK- 696
Db 2846 SFDNNFLISGQCGQCONSFYKVMAGTEYVDKALLATGCGNV---APGVYITPSNFTAV 2901
Qy 697 -KTNKGPVAGIY-----NGNINTEMPROPKFKQGRKDNDSHYAVIED-- 739
Db 2902 FQGOEAPAGCFASFSVRCGSNFTGPSYIIS--PMYPRQY-----DNMNCNTYIIEANP 2954
Qy 740 -----TWYGHLLQDS--SGSLQPEVDYTR-----PFGQTMGVCPPSPPTI 779
Db 2955 LSVYLLTFVFSFHLARSAYTGSVNDGVHIIIRQYSVMSPTFATVCGDEMPAPLTI 3009
```

RESULT 5

S58984
development protein tolkin (EC 3.4.24.-) - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 19-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 21-Jul-2000
C:Accession: S58984
R:Pinelli, A.L.; Xie, T.; Bossie, C.A.; Blackman, R.K.; Padgett, R.W.
Genetics 141, 271-281, 1995
A:Title: The tolkin gene is a tollold/BMP-1 homologue that is essential for Drosophila C
A:Reference number: S58984; MUID:96042912; PMID:8536976
A:Accession: S58984
A:Molecule type: mRNA
A:Residues: 1-1464 <FIN>
A:Cross-references: EMBL:U04777; NID:g1002985; PIDN:AAC47015.1; PID:g1002986
A:Note: the authors did not translate the codon for residue 722
C:Genetics:
A:Gene: tolkin
A:Cross-references: FlyBase:FBgn0004885
C:Superfamily: astacin homology; EGF homology
C:Keywords: hydrolase; metalloproteinase; zinc
F:529-722/Domain: astacin homology <AST>
F:958-993/Domain: EGF homology <EGF>
F:118-1153/Domain: EGF homology <EGF1>
F:614,618,624,673/Binding site: zinc (His, His, His, Tyr) #status predicted
F:615/Active site: Glu #status predicted

Query Match 2.7%; Score 117.5; DB 2; Length 1464;
Best Local Similarity 19.3%; Pred. No. 5.5;
Matches 98; Conservative 63; Mismatches 161; Indels 185; Gaps 24;

QY 105 PFGVYQLPSTSLPTNRFIVDVKAHKSIGLEQSIIRLRKIGGSCPGCVTSTIS 164
DB 1004 PNGIT-TSPFPPEWYPLKCKCIWETVAPPKHRIISLNTFDL-----EETAHQQS 1052
QY 165 --GRIDATVY-----RIGFECNGTVSRIKMGVKKALHPVHRNV--SGFSI 211
DB 1053 DCGVDSTVYSKIGENLKIKIGFCGSSIRPTATSEN--ALRLPHSDKSIORSFPA 1109
QY 212 ANNSIKR-----LCIIIEYF-----EGEGS-----ATL 235
DB 1110 VFFDIDIECAVNNCGCCECRNTIGSYICMCHNGYSMHENGHDCGECKYEISAPRGTI 1169
QY 236 MSNYPGPFDELMTQGFV-VRPHLRASVFLNPL-SNCERKEEVEVEYIIPSTINPE 293
DB 1170 FSPYRPSTYPPNDCVNHFTTTPGH-RIKLIFNEDESIOECTYDVNAVY----- 1219
QY 294 VFLEKQPKPMAGNENLSIQGCDQDAQSPGIL-----RLQFQVLYQHNPESNKIYV 347
DB 1220 -----DGESESSVYLRFCGDKIIPPI-----SSISNOMYV 1251
QY 348 DLSNERAMSLTIEPRYKQRKVEYPCVLESFRTGSSNLTLSGSKHKISFL-----C 401
DB 1252 -----LKTQKNKOKNGFTASHSPACGGLFRATSOVQOFYSHAFGNODY 1295
QY 402 DDLRLMANNVEKTICTGDHYCO-----RKYS-----LQVPSILH----- 438
DB 1296 DD-----GMDEWITIAADNSYVQLIFLPDISSENCTFDYVOVFSIDIDVGYGPMYG 1351
QY 439 -----LPVELHDFSWKL-----VPRDLRLVLPVPAQKLOQHTH-EKPCNTSEYL 483
DB 1352 QYCGNVLPQDINSWTHSLVLRPKDGSVPKGFASVVAAPNNGSEYHSDEDVENSISSE 1411
QY 484 VASAIPIQDLYFGSFCGSGSIKQIQVK 510
DB 1412 MVTPEP-----GSLKSIYIE 1426

RESULT 6
T24315
hypothetical protein T01G9.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T24315
R:Lennard, N.

submitted to the EMBL Data Library, July 1996
A:Reference number: Z19874
A:Accession: T24315
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1603 <MID>
A:Cross-references: EMBL:Z19874; PIDN:CAB00050.1; GSPDB:GN00019; CESP:T01G9.3
A:Experimental source: clone T01G9
C:Genetics:
A:Gene: CESP:T01G9.3
A:Map position: 1
A:Introns: 68/3; 346/1; 407/3; 450/3; 486/3

Query Match 2.7%; Score 117; DB 2; Length 603;
Best Local Similarity 19.9%; Pred. No. 1.6;
Matches 133; Conservative 79; Mismatches 216; Indels 242; Gaps 32;

QY 266 FLNF-----NLSNCKEKREVEYIIPGSTNPEVF-----KLEDK 300
DB 9 FLSPFLVCSIALPSSCPNLCECDQNDSSWSYCKRAIINDIYAEILNOLPLTLRLSHIQ 68
QY 301 QPGMAG-----NFN-----LSIQGCDQDAQSPGILRLQFOVLYVHPON-----E 340
DB 69 PPSNRIGSNKLRNNDINRPAQLVRLINCQIPAMRSIRLSLEVLDLHNNIEHATM 128
QY 341 SN-----KIYVDLSNERAMSLTIEPRYKQRKVEYPCVLES-RTGS-SNLTITSGS 393
DB 129 SNFGQMKRLVLDSSNH-----LNLTP-----TGVFTYLALRLSLSNNTSIDLS 175
QY 394 KHKISFL-----CDDLRLMANNVEKTICTDHYR---QQRKS-YSLQVPSD 435
DB 176 TNLRLGINSRLVRLDRNPRIPIEIHINELFTDY-----SDLELYLHNCILSTISLADR- 230
QY 436 ILHLPELHDFSWKLVPKDRLL-----SLVLPAPKLOQHTHERPCNTSESYLVASAP 489
DB 231 -----IPQLRLQIGIGNNLKMPRTKELRS-----LPQLSVLDLSHNS 267
QY 490 SQDLYFGSFCPGSGSIKQIQVKONT-----SVTLRTFAPSF----- 524
DB 268 IOETTACAF-C-NFNIISKLDLSHNLGISKDSPFNEDAFRTMPRLRLDLFSFHNHNDPDSKW 326
QY 525 -----QQRASQGLVSTIPIPFKEGVFTYPRDKSKYLLTP-NMDGLP----- 569
DB 327 LGMAQELTSLIALSGNLKMFESWYTL---KSLHLRLAYAHKIFIPQLPSRYHL 382
QY 570 -----SLTSYMNISYPRDQVACLPFKERGVVCOGRAPMIIQEQ----- 611
DB 383 ISLNISSNELTYLPDINILTPNVTITANRPHRTSHTDLFLNNEQVYVYDGNPWC 442
QY 612 -----RTAAELTSLDEVDLPKPSF---HHHSFWNISNCSPTSGKQDLDF 655
DB 443 SCALQGLQVHMRDRYARHILNVDNVCAPSLVEGSHVLAITDVNCAVLFGARGL- 500
QY 656 SVTLTPRTVQLVYLIAVGGVILLALDLIC-CYKKKKKTKNRPAYGIYGNINTE 714
DB 501 -----TQTSMLLIL-----AGVLLFAALLMILGICITPLREKQYK-----SYTYRE 543
QY 715 MPRQPKFKQGRKDNDSHYAVLEDTVVYUHLQDSSGSF---LQPEVDYV---RPFQ 767
DB 544 HSKRPL-----TMANHSCSSSTINDHGLPSLPFPDFLVSTEFKA 584
QY 768 TMGVCPSPSP 777
DB 585 TPPLIPAPP 594

RESULT 7
T34022
zonadhesin - pig
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C:Accession: T34022
R:Hardy, D.M.; Garbers, D.L.

A:Residues: 1-582 <MIL>
A:Experimental source: strain A
A:Note: Strain A is a helper virus of the strain T
C:Comment: Enzymatic cleavages of env polypeptide may yield mature proteins including co
C:Genetics:
A:Gene: env
C:Superfamily: type C retrovirus env polypeptide
C:Keywords: coat protein; polypeptide

Query Match 2.6%; Score 115; DB 1; Length 582;
Best Local Similarity 19.1%; Pred. No. 2.1;
Matches 89; Conservative 60; Mismatches 147; Indels 170; Gaps 19;

QY 314 QCCDQADQSGEILRLQFVL-VQHPQNSKTIYVDL-----SNERASLTIEPRPVK 365
DB 221 RGVLDLPDSTLDEATQVFNATNPQLENCMLCMTLTGTSPPSRMAMSLME----- 275
QY 366 QSRKRVGCVCELSRSCSSNLTLTSGSKHIS--FLC---DDLTLMMVENTISCTD 419
DB 276 -----LAVLASLSGATHRVNRCQLLCRADNRGTIPVGYVHTCTST 317
QY 420 HRYCORSYSLOYPSD-----LHLPELHDFSWKLLPXRDLSTLVVPAOKLQOHTHEK 474
DB 318 IQESLTRVYIELRDVYVLRVWVLCVEQHAYT--ALPKWIGL----- 359
QY 475 PCNTSESYLVASALPESDLYFGSPGSGIKQIQVKNISVTLTFAPSFOQERSKGLT 534
DB 360 -C-----ILASIVPDM-----SIIPG-----EPIPL-----PSLEYTAGRHKRA 393
QY 535 VSFIPIKESGVFTV-----PDRKSK 556
DB 394 VQFIRPLVGLTGATAGTGLGVSVHYHKLSNQLIEDVALSGINDLQDIDSLAE 453
QY 557 VYLTPPMWDRGLSLTSVSNISVPRDOVACLTFPKERSGVVCTGRFMIQORTRAE 616
DB 454 VVLON---RKGDLTLHBOGICLALOEKCC--PYANKSGV-----RD 492
QY 617 EITSLDEDVLR-KPSFHNSFWNISNCSPTSGKQDLLESVTLTPRTVDTVLIAVAG 675
DB 493 KIRKIQEDLLARKRALYDNLNGLNGLFLPYLLPSLGLP----- 532
QY 676 GGVLLSALGLIICCVKKKKTKKGPVAVGYNGNINTEMPRORPK 721
DB 533 -GLTLFLTLG---PCIKRTLTIRIHKIQGSKNPRISPAVQAPR 574

RESULT 10

S46754
hypothetical protein YHR155w - yeast (Saccharomyces cerevisiae)

C:Species: Saccharomyces cerevisiae
C:Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 19-Apr-2002
C:Accession: S46754
R:Macrl, C.
submitted to the EMBL Data Library, June 1994

A:Description: The sequence of S. cerevisiae cosmid 9666.
A:Reference number: S46752

A:Accession: S46754
A:Molecule type: DNA

A:Residues: 1-1228 <MAC>
A:Cross-references: EMBL:U010397; NID:9500647; PID:9500650; GSPDB:GN00008; MIPS:YHR155w

C:Genetics:
A:Gene: MIPS:YHR155w
A:Cross-references: SGD:S0001198
A:Map position: 8R

C:Superfamily: pleckstrin repeat homology
Query Match 2.6%; Score 114; DB 2; Length 1228;
Best Local Similarity 18.8%; Pred. No. 7.6;
Matches 99; Conservative 85; Mismatches 196; Indels 148; Gaps 21;

QY 179 SNGVSVIRIKQEGYKMAHLPMFPRNVSGFSTANRSSIRLCIES---VPEGESAT 234
DB 301 SNGNSEEKGLSG-----WLYMKTTVGHDPKRVVVRMCFLQNNVGFVSLSPSKT 352

QY 235 LMSANYPEGPEDELMTQFVVPVRAHLASVFLNLSNCRKEEREVEYIIGSTTNEV 294
DB 353 YVEETDFKG-----ILMTIV-----EYLP----- 371
QY 295 FKLEDRQPGMAGNFNLSLOGCDDAOSPGLRLQFQVIVQHPQNSKTIYVDLSNERA 354
DB 372 ----KEPRNCEKRLRIQNPCKTEENTYI---DIIQAESIDELKSWINTLTSKR 421
QY 355 MSLT-----EPRVYKORKEVPGCVLESRTCSNLTLSGSKHISFLCDDTLRLMN 410
DB 422 IALSIKEENPRYLARKKIEPQFEFASSSTDKLTLFSSTKTL-LVEELKKYMS 480
QY 411 VEKTSCTHRYCORSKYSLOVPSDILHPELHDFSWKLLVPR----- 454
DB 481 EDDIYSITIDMKAYHLRIVSPIATQTLHLA---FSTFLSVSNVYPCATQANTWTAMW 536
QY 455 DRSLVLVPAOKLQOHTHEKRCNTPSFYLVASALPS-----QDLFGSFCGGSGIKQ 506
DB 537 NDLSYLVNPLKGSVH---KPAVYNSRSRVSIPDYPSLKVDIDQFRSIFSVNDF 593
QY 507 IOVKONISVTLR--TPAPSFQDE-ASRQGLTVSFPIYFKEEGVFTVTP-----DTRSKV 557
DB 594 LQVPEL-VLLRYSVWCNPKKQKFAFMAFYLHITVYLINISGFSYLRLDLDISIE 652
QY 558 YLRTP-----NMDRGLSLTSVSNISV-----PRDOVA 586
DB 653 YDKSPKRVSSRMLMQG---DGLRFMSVFTDRBRAVASKLOFLIENKAMHPIKGEKE 708
QY 587 CLTFPKERSGVVCTGRFMIQORTRAEITSLDEDVLRKPSFHNN 634
DB 709 VLEIFQELDEIENEKR--IKDNLSESH-YSKDYDLLKSTYDHH 752

RESULT 11

T40765
web1 protein homolog - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T40765
R:lyne, M.; Rajadream, M.A.; Barrell, B.G.; Lauber, J.; Hilbert, H.; Dueterhoeft, A

submitted to the EMBL Data Library, March 1998
A:Reference number: T40765
A:Accession: T40765
A:Molecule type: DNA

A:Status: preliminary; translated from GB/EMBL/DDJ
A:Cross-references: EMBL:AL022072; PID:CAA17835.1; GSPDB:GN00067; SPDB:SPBC8D2.20C

A:Experimental source: strain 972h-; cosmid c8D2
C:Genetics:
A:Gene: SPDB:SPBC8D2.20C
A:Map position: 2

Query Match 2.6%; Score 113; DB 2; Length 1224;
Best Local Similarity 17.9%; Pred. No. 9;
Matches 153; Conservative 108; Mismatches 372; Indels 222; Gaps 32;
QY 123 RTFIMDYKAKHSIGLELOFSIPRLQIGPES-CPDG-----VTHSISGRID----- 168
DB 284 RAMWVNVETGESLG-----SFPKSGNMVTKSSWCPSSNRVAVASLGGKVSIFSIQSTNT 338
QY 169 -----ATVVRIGTFCSGNGTVSRIKQEGYKMAHLPMF-----HPRVSGF 209
DB 339 DKQSEASIKGATSIDNDEFNNLPSIGSDEPSFLPLAPKMFVYPVARGPKNKIYVF 398
QY 210 SIANRSSIKRLCITESVFEDEGSATLMSANV-----PGCFPEDELMTQFV 255
DB 399 S-PSKSEVLTTSAPDEVEDQDEAKSFHSSAKFQTEKETIDPCQKGVESASEEPAIMNKL 457
QY 256 VPAHLRAS-----VSFLNPNLSNCRKEEREVEYIIGSTTNEVFKLEDRQPGMAGN 308
DB 458 MAVSKRASRSKFAELLGYKTLKPRNDEDSKVDSEVAKDSTTPNELSK----- 505

```

Oy 309 FNLISGGDDDOAQSPGILRLQFO--VLVOHPON-----ESNKIYVVDLSN--ERASLITEP 361
Db 506 -NANNENTDDSSSYGKLAEVSQEVSTADCKDAELVADSFITFPEDSDELKNTTEALLT 564
Oy 362 RPVKOSRKFVPGCEVCCESRTCCSSNLT-LTSGSKHKISFLDDDLRLMANNKETSCTDH 420
Db 565 GDVLISAVK-----ACDEKKISEALFLSTFGKNE-----CRKCYR-----DARYEIQEH 608
Oy 421 RYCQRKSTSLQVPSDILHLPELHDFSKW-----LLVKKRSLSLVLPQAQKIQDHTH 472
Db 609 KPSYMRSLACIADNDLQVVDNAEVSWEKDDIFVEICATATDEEAPRLCSTGLQRELDLED 668
Oy 473 EKPQNTS--FSTLVVSAIPS--ODLYFGSPCGSGIKOIQOVKNISVTULTFAPSQOEA-- 528
Db 669 EKSTRSAPFCITASKSLDSIYANLW-----LKQLATSTKTSKAASAYCAIYEQLTKL 719
Oy 529 -SRQGLTVSFIPYERKEGVFTVPDPTK-----SKYLLRTPMMDRLPLSLTSYSW 576
Db 720 MDKVMFSRIYVYKDE--LSATKDKMKGALYEYIAYAKLLSASGRKDDAMSYLNLVPT 777
Oy 577 NISVPDROVALTLFFKESGV-----VCQTG-----RAFMIQEQR 612
Db 778 EPPGAKEEIQRLTMYLLENAVAPRIHQIKQTSYAPVQPKTSQASSILPTVPRTSYTSEYA 837
Oy 613 TRAEIFSLDELDVLPKPSFHHHSFW-----VNISNCSPST 647
Db 838 TTSSTHTPADVHPLRPSTSTTAGNDAPMLGQLEPMRAAASMAPVRSPPFGASSAPDA 897
Oy 648 GKQDLLEFSVTLTPRTYDVLVILIA-----AVGGVLLLSALGLLICVKKKKK-- 697
Db 898 MSRTSSVSTLPPPTASMTASAPALASPPPKVGETYHPPTAGTRVPVQOPSHMPY 957
Oy 698 TNKGPAVGIYNGNINTEMRQPKFKQGRKNDNSIVAVIEDTWYGHLLDDSSGSFLQP 757
Db 958 TPVAPQSPVAASARISSSPNMP-----PSNPYTPPIAVASSIVNPAH----- 998
Oy 758 EVDVTRPFGQTMGVCPPSPPTICSHAPYAKLATEEPRPSPSESEBPTFSHPNCGDVS 817
Db 999 ---TYKPIGSGQIYPPRQP-----ANRVVPLPPTASGRASAYEPPTYSVSPSALS 1047
Oy 818 SKDR-DIPLANTQEP 831
Db 1048 PSYTPQLPVPVSSRLP 1062

RESULT 12
T42205
breast cancer susceptibility protein BRCA2 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C:Accession: T42205
R:McAllister, K.A.; Haugen-Strano, A.; Hagevik, S.; Collins, N.K.; Brownlee, H.; Futreal,
submitted to the EMBL Data Library, February 1997
A:Reference number: 222073
A:Accession: T42205
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-3329 <MCA>
A:Cross-references: EMBL:U089652; NID:92443438; PID:92443439; PIDN:AAB71377.1
A:Experimental source: strain CD1; 129Sv; ICR Swiss
C:Genetics:
A:Gene: BRCA2
C:Superfamily: breast cancer tumor suppressor BRCA2

Query Match 2.68; Score 113; DB 2; Length 3329;
Best Local Similarity 18.88; Pred. No. 41;
Matches 125; Conservative 94; Mismatches 226; Indels 220; Gaps 33;
Oy 198 LPWHPRNVGSF-----SIANRSSTIKRLCTIESVREGSGSATMSANPDEGPEDEL 249
Db 1188 LPNINEMEGGFCALGTLKLSVSNDEALKRAKMLFSDIENSEEPSAKVG--PRGFSSS-- 1242

```

Oy 250 MTWQVVAHLRASVSLFNENLNCER-----KEEREYIIPDS---TT-----NEVVK 296
 Db 1243 -----AHHDVASVAFKIKKONTEKSFDEKSSKQVTLONNIEMTICIEVGRANEKTI 1294
 Oy 297 LEDKOPGNMAGNFNLSLQCDQDQDQSPGILRLQFVLVQHPONESKRIIV---VDLSNE 352
 Db 1295 KNTKHEDDVTSQGRNNLEMSDGSMS-----TSGPVYIHKGDSDL----- 1334
 Oy 353 RAMSLTIEPRPVKQSRKFVPGCFVCLLESRTGSSNLTLTSGSKHKISFLCDLITRL-WMNV 411
 Db 1335 -----PADQSK-----CPESCTQYAREENFTQ-IKENIS-----DLTCLTEIMKA 1372
 Oy 412 EKTICTDHRKYQQRKSYSLQVPSDLHLPEVLDHPSW----- 448
 Db 1373 EET-----CMKSSDKQLPSPDKMEONIKIEFNISFTQASGNKTRVSKESLKNYSNIEN 1424
 Oy 449 ----KLVPKDRU-SLVLYVPAOKLOOH--HEKPCMTSPSYLVASAIPODLVYFGSFCPG 501
 Db 1425 RETDELVTYISDBLSNKLHLGINKDKMHTSCHK-----AISTKKYFEDHF-PI 1471
 Oy 502 GSIKOIQVKONISVTLRTFPAPSFQOEASHOGLTVSE-----IPYFKE--EGVFTVTPD 552
 Db 1472 VTVSQPLAQOH-----PEYEIESTKEPTLLSPHTASGKKVKIMQSLDLKVKLJFDE 1522
 Oy 553 TKSXYLR-TFPMWDGCLPBLTYSVSNISVPROVACLTFFKERSGVCOGTGRFMIIQED 611
 Db 1523 TQ--YVRKTAFSQSGSKPL-----KSKKELTAYEK-----IEVT 1556
 Oy 612 RTRAEEI--FSLDDVLPRPSFHHSFWVNISNCSPTSGK-OLDLFSLVTLPRTVLTL 667
 Db 1557 ASKCEMGNFVSKETEMLPQONTNHTROTENLKTSGTSSKYQENIENNVEKRP----- 1611
 Oy 668 VILIAAGGVLLLSALGLIIICVKKKKRKTNGPAVGIYNGNINTEMRQPKFQGRK 727
 Db 1612 -----ICICQSSYPVTEDSLALAYTEDSRKTCVRE--SLSKGRK 1650
 Oy 728 -----DNDSHYVAVIDEDTMY-GHLLQDSSGSFLQPEVDITRRPFGCMGVCPPS 775
 Db 1651 WLREGDKLGTNTIKIECVKHEHTEDFAGNASYEHSLVIRREIDNHVSENOVSTL-LS 1709
 Oy 776 PTTC 780
 Db 1710 DPNV 1714

RESULT 13
 T30904
 breast cancer tumor suppressor Brca2 - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jan-2000
 C:Accession: T30904
 R:Connor, F.; Smith, A.; Wooster, R.; Stratton, M.; Dixon, A.; Campbell, E.; Tait, T
 Hum. Mol. Genet. 6, 291-300, 1997
 A:Title: Cloning, chromosomal mapping and expression pattern of the mouse Brca2 gene
 A:Reference number: Z20931; MUID: 97217789; PMID: 9063750
 A:Accession: T30904
 A:Status: preliminary;
 A:Molecule type: mRNA
 A:Residues: 1-3329 <CON>
 A:Cross-references: EMBL:U82270; NID:g1854950; PID:g1854951; PIDN:AA848306.1
 C:Genetics:
 A:Gene: Brca2
 A:Map position: 5
 C:Superfamily: breast cancer tumor suppressor BRCA2
 C:Keywords: tumor suppressor

Query Match 2.6%; Score 113; DB 2; Length 3329;
 Best Local Similarity 18.8%; Pred. No. 41;
 Matches 125; Conservative 94; Mismatches 226; Indels 220; Gaps 33;

Oy 198 LPWFPRNVSGF-----SIANSSIKRLCIIIESVFEGESATILMSANYEDGPEDEL 249
 Db 1188 LPNIMEFEGGCSALGTRKLSVSNALRKAMKLFSDIENSEPSSAKVG---PRGFS-- 1242

This Page Blank (uspio)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 11, 2002, 05:03:00 ; Search time 21 Seconds

(without alignments)
1651.153 Million cell updates/sec

Title: US-09-899-569A-4
Sequence: 1 MAGLNCGVSLALGVLLGA.....SSKNDIPDLMTQPEMPDAE 836

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SWISSPROT_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	118	2.7	2499	1 MPRI_BOVIN	P08169 bos taurus
2	117	2.7	2476	1 ZAN_PIG	Q28983 sus scrofa
3	115.5	2.6	849	1 PER_ANTPE	Q17062 antheraea p
4	115	2.6	582	1 ENV_AVIRE	P03399 avian retic
5	114	2.6	1328	1 YHV5_YEAST	P38851 saccharomyc
6	113	2.6	3329	1 BRC2_MOUSE	P97929 mus musculu
7	112.5	2.6	440	1 OMGP_HUMAN	P23515 homo sapien
8	112.5	2.6	818	1 PTR2_YEAST	P47116 saccharomyc
9	112.5	2.6	1805	1 RML_HUMAN	O92545 homo sapien
10	112	2.5	1694	1 SN_MOUSE	O62330 mus musculu
11	111	2.5	794	1 TRKL_LYMS	O76897 lymnaea sta
12	110.5	2.5	704	1 CRAR_MOUSE	P98064 mus musculu
13	110.5	2.5	942	1 TMKL_ARATH	P43398 arabidopsis
14	109.5	2.5	692	1 DPE2_YEAST	P24482 saccharomyc
15	108.5	2.5	1828	1 MAP2_MOUSE	P20357 mus musculu
16	108	2.5	699	1 CRAR_HUMAN	P48740 h complem
17	108	2.5	838	1 SMAG_HUMAN	O98099 homo sapien
18	106	2.4	777	1 UNC8_MOUSE	O21574 caenorabd1
19	106	2.4	830	1 YBMA_SCHPO	O10332 schizosacch
20	106	2.4	1091	1 NCAL_CHICK	P13590 gallus gall
21	105.5	2.4	913	1 CAD4_CHICK	P24503 gallus gall
22	105.5	2.4	4590	1 FATH_HUMAN	O14517 homo sapien
23	104.5	2.4	810	1 ILAR_MOUSE	P16382 mus musculu
24	104.5	2.4	1490	1 CRK7_HUMAN	O98944 homo sapien
25	104	2.4	449	1 PCOL_HUMAN	O15113 homo sapien
26	104	2.4	4036	1 RRP1_DUGBY	O66331 dugbe virus
27	103.5	2.4	727	1 CTCE_HUMAN	P48922 homo sapien
28	103	2.3	2347	1 KROS_HUMAN	P08922 homo sapien
29	103	2.3	2700	1 ZAN_HUMAN	O94493 homo sapien
30	103	2.3	2768	1 THRG_HUMAN	P01366 homo sapien
31	102.5	2.3	457	1 CD4_MOUSE	O29037 salmistr sci
32	102.5	2.3	457	1 CD4_SAISC	P24063 mus musculu
33	102.5	2.3	1163	1 ITAL_MOUSE	

ALIGNMENTS

```

RESULT 1
MPRI_BOVIN
ID MPRI_BOVIN STANDARD: PRT: 2499 AA.
AC P08169;
DT 01-AUG-1988 (Rel. 08, Created)
DI 01-OCT-1989 (Rel. 12, Last sequence update)
DI 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cation-independent mannose-6-phosphate receptor precursor (CI Man-6-P
DE receptor) (CI-MPR) (Insulin-like growth factor II receptor) (300 kDa
DE mannose 6-phosphate receptor) (MPR 300) (MPR300).
GN IGF2R OR M6P.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_Taxid:9913;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE-8115411; PubMed-2963004;
RA Lobel P., Dahms N.M., Kornfeld S.;
RT "Cloning and sequence analysis of the cation-independent mannose 6-
RT phosphate receptor."
RT J. Biol. Chem. 263:2563-2570(1988).
[2]
RP SEQUENCE FROM N.A.
RX Killian J.K.;
RA Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE OF 1039-2499 FROM N.A.
RX MEDLINE-87175648; PubMed-2951738;
RA Lobel P., Dahms N.M., Breitmeyer J., Chirgwin J.M., Kornfeld S.;
RT "Cloning of the bovine 215-kDa cation-independent mannose 6-phosphate
RT receptor."
RT Proc. Natl. Acad. Sci. U.S.A. 84:2233-2237(1987).
-1- FUNCTION: TRANSPORT OF PHOSPHORYLATED LYSOSOMAL ENZYMES FROM
CC THE GOLGI COMPLEX AND THE CELL SURFACE TO LYSOSOMES. LYSOSOMAL
CC ENZYMES BEARING PHOSPHOMANNOSYL RESIDUES BIND SPECIFICALLY TO
CC MANNOSE-6-PHOSPHATE RECEPTORS IN THE GOLGI APPARATUS AND THE
CC RESULTING RECEPTOR-LIGAND COMPLEX IS TRANSPORTED TO AN ACIDIC
CC PRELYSOSOMAL COMPARTMENT WHERE THE LOW PH MEDIATES THE DISSOCIATION
CC OF THE COMPLEX. THIS RECEPTOR ALSO BINDS INSULIN GROWTH FACTOR II.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Lysosomal.
CC -1- DOMAIN: CONTAINS 15 REPEATING UNITS OF APPROXIMATELY 147 AA. THE
CC MOST HIGHLY CONSERVED REGION WITHIN THE REPEAT CONSISTS OF A
CC STRETCH OF 13 AA THAT CONTAINS CYSTEINES AT BOTH ENDS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sib.ch).
CC
CC EMBL: J03527; AAA30455.1; -
CC EMBL: AF342811; AAL23908.1; -

```

```
DR PIR; A30788; A30788.
DR HSSP; P02751; 2FN2.
DR InterPro; IPR000479; CIMR.
DR InterPro; IPR000562; FN_Type_1I.
DR Pfam; PF00040; fn2; 1.
DR Pfam; PF00878; CIMR; 13.
DR PRINTS; PR00013; FNTYPE1I.
DR PRODOM; PD000995; FN_Type_1I; 1.
DR SMART; SM00059; FN2; 1.
DR PROSITE; PS00023; FIBRONECTIN_2; 1.
KW Transmembrane; Transport; Glycoprotein; Repeat; Receptor; Lysosome;
KW Signal.
FT SIGNAL 1 44
FT CHAIN 45 2499
FT
FT DOMAIN 45 2313
FT TEANSEM 2314 2336
FT DOMAIN 2337 2499
FT REPEAT 45 170
FT REPEAT 171 327
FT REPEAT 328 478
FT REPEAT 479 629
FT REPEAT 630 771
FT REPEAT 772 933
FT REPEAT 934 1089
FT REPEAT 1090 1229
FT REPEAT 1230 1373
FT REPEAT 1374 1518
FT REPEAT 1519 1658
FT REPEAT 1659 1807
FT REPEAT 1808 1999
FT REPEAT 2000 2137
FT REPEAT 2138 2290
FT DOMAIN 1909 1951
FT CARBOHYD 120 120
FT CARBOHYD 409 409
FT CARBOHYD 444 444
FT CARBOHYD 552 552
FT CARBOHYD 590 590
FT CARBOHYD 635 635
FT CARBOHYD 755 755
FT CARBOHYD 879 879
FT CARBOHYD 959 959
FT CARBOHYD 1030 1030
FT CARBOHYD 1173 1173
FT CARBOHYD 1255 1255
FT CARBOHYD 1321 1321
FT CARBOHYD 1665 1665
FT CARBOHYD 1766 1766
FT CARBOHYD 1825 1825
FT CARBOHYD 2094 2094
FT CARBOHYD 2145 2145
FT CARBOHYD 2220 2220
SQ SEQUENCE 2499 AA; 274526 MW; 3C1C9DEF2875159D CRC64;

Query Match 2.78; Score 118; DB 1; Length 2499;
Best Local Similarity 19.98; Pred. No. 3.2;
Matches 195; Conservative 100; Mismatches 343; Indels 340; Gaps 56;

QY 1 MAGLNCVSTALLGVLLGGAARLRGAAPFIALPRESNTVLKIGTPILL---AKPC 56
DB 1710 MHGLACPAGTAVCKVPADGP-----PIDIGRAVAPPLINPLANEVLYNESSTPC 1759
QY 57 YIVSKRHITLISIKSGERIVTFSCQ-----SPE-----NHFVIEIQKNIDC---M 100
DB 1760 ---LADRHENTSL-----ITFHCKRGVSMGTPKLLRTSVCFVEMWEPVLCPPREVK 1809
QY 101 SGCPPEFGEVOLQPSLPLPLNRTFIMDVAAHK-SIGLEIQFSLPRLRQJPGG-SCPDG 158
DB 1810 TDGSLSLDEQLDYGFNLSLSKSTFKYTRGPHYTSVG-----CTAAGLDEGGCKDG 1862
QY 159 VTHSISGRIDATVVRIGTFCSNGTVSRIKKQEGKAMALHLPWFPRKRVSGSINRBSIK 218
```

```
DB 1863 AVCLSSSKASRGRLA-----SMKLDYRHQDAVLLSTANGTCF 1903
QY 219 -----RLCIIEVFEEGSATLMSANYPGFPEDELMTOFVPAHLRASVSPLNLS 272
DB 1904 PETEGEPCVFPFVFENK-----SYECVYSRARLW-----CATPANDRD 1945
QY 273 N---CERKEERYEYIIPGSTNPEVKLEDKQGNMAGNFNLSLQCCDDAOSGILRL 328
DB 1946 HEMGFCKHSTSH-----RTSVILPK-----CDEAD----- 1971
QY 329 QFYVLVHPNENSKKIYVVDLSNERAMSLTIEPR-----PVKSRKFV----- 371
DB 1972 -----VRPO-----VFSEVRGCEVFEMTKKVCPPKKKECFVQKHRTYDLRL 2017
QY 372 ---PGCF-----VCLF-----SRTCSSNLTL-----TSGSKRISF-----LC 401
DB 2018 SSLGSMSPFNHGASYYINLQKIKYKGPQDSERASYSCKSTSEGVGLVHTQKLDVY 2077
QY 402 DDLTRLMMANVEKTICTDHRVQ-----RKSYS-LOYPSDILHLPVLDHPSW 448
DB 2078 DD--RVIVTYSKGHYCGDNKTASAVIELCAKTVGRSPSTFRVDSCSTYH-----FSW 2128
QY 449 KLVPRKRLSLVLPAAKLO--QHTHEKPCN-TFSYLVASAIKPSODLYGSPGGSIK 505
DB 2129 -----DSRAACAVKPOEVQKNGVTITNPANGSFSL-----GDIYKRFSASGDVR 2174
QY 506 -----QIOVK-----ONISVTLRTFAPSFOEASROGLTVSFLPYKEBG-- 545
DB 2175 TNGDRYIEIQLSSITGSSSPACGASICQK--KANDQHSRKRVGTSNQRYYVQGD 2231
QY 546 --VFT-----VTPTKSKVYLR-TPNMDRGLPSLTSWMSNLSVRDQVACLTFFKER 594
DB 2232 DVFTSSSKGCKDKTKSVSSTIFFHCDPLVKGDIPEFS-----HETADCYLFSWH 2282
QY 595 SGVVCQGRAFMILQEQRTAEIFSLDEVDLPPSPHHSHFWNISNCSPTSQKDL 654
DB 2283 TSANCPGAGF-----DETAGDAQDEHK-----GLSERQAVAVNLSL 2322
QY 655 FSVTLPRPTVDLVIILAAVGGVLLSALGLIIC-----VKKKKKTKNGPVGITYNG 709
DB 2323 ---LVALTGLCLLTLVLLKKERREVMVSR--LTNCCRANSAVSKYSKVKEEBA---D 2372
QY 710 NINTE-----MPQPKKFKQGRKNDSHYAV---IEDTM--VYGLLDSSSFLQ-P 757
DB 2373 ENTEWLMEEIOPAPRPGEGOE--NGHVAAKSVRADTLTSLAHG--EDDSDEVLTLR 2429
QY 758 EVDIYRPFQ--GTGVCPPSPPTCSRAPTAK-----LATEBPSPSPSESEPTFS 809
DB 2430 EVKVRPPGRAPGAG--GPLRLPLPRKAPPLRADRVGLVRGEPRARRGPRAAATPISTF 2488
QY 810 HPNNGDVSSKDTDIPILN 827
DB 2489 H-----DSDDEDLH 2498

RESULT 2
ZAN_PIG ID ZAN_PIG STANDARD; PRT; 2476 AA.
AC 028983;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Zonadhesin precursor.
GN ZAN.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 823-830; 859-872; 883-890;
RP 920-925; 960-967; 1235-1244; 1349-1354; 1518-1532; 1624-1656;
RP 1658-1667; 1777-1795 AND 1914-1921.
RC STRAIN=Meishan; TISSUE=Testis;
```


CC -1- MISCELLANEOUS: STRAIN A IS A HELPER VIRUS OF THE STRAIN T.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: K02537; AAA9198.1; -
CC EMBL: X01455; CAA25686.1; -
CC PIR: A03999; VCDAR.
CC HSSP: P03385; 1MOR.
CC InterPro: IPR002050; Env_polyprotein.
CC Pfam: PF00429; Env_polyprotein.1.
CC Coar protein; Glycoprotein; Transmembrane; Polyprotein.
CC CHAIN 1 391
CC FT CHAIN 1 391
CC FT CARBOHYD 241 241 COAT PROTEIN GP72.
CC FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 314 314 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 485 485 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SEQUENCE 582 AA; 64138 MW; CD2560ADF026D32 CRC64;
Query Match
Best Local Similarity 2.6%; Score 115; DB 1; Length 582;
Matches 89; Conservative 60; Mismatches 147; Indels 170; Gaps 19;
OY 314 OGCDQDASPGILRLQFOVL-VQHPONESNKIVVDL-----SNERAMSLTEPRPVK 365
DB 221 RGVLDLPQSTDLIETHQVLAATNPNQLAENCMCLTLGTSQPSRPMASLSME----- 275
OY 366 QSKRPFGCFVLESPTCSNLTLTGSKHKIS---FLC---DDLRLMKNVEKTSCTD 419
DB 276 -----IAVLASLSGATHRVNRCOLLCREADNRGTGIPGVYHFTNCTS 317
OY 420 HRYCOKRSYSLOYPSD-----ILHLPELHDFSMKLLVPRDRSLVLAPOKQLOHTHEK 474
DB 318 IQSSLTRRYVYELRDVYLHRYVILCEVQHAYT---ALPKMKIGL----- 359
OY 475 PCMTSFYSYVANAIPSDQLYFGSCFPGSGSIKIQVONKONISVTLRTFAPSFOEASRQGLT 534
DB 360 -C-----ILASIVPDM-----SLTPG-----EERIPL-----PSLEYTAGNKKRA 393
OY 535 VSTIPFKEGEVFTV-----PRTKSK 556
DB 394 VQPIPLVGLIGATLGGTGLGVSVHTYHKLNSQLIEDVQALSGTINDLOQIDSLAE 453
OY 557 VYLRTNMMDGRLPSVSNINISVPRDQVACLFFPKERSGVNCGTGRAFIIDQORTRAE 616
DB 454 VVLQN---RRGLDLTLAEQGGICLALQEKCC-FYANKSGIV-----RD 492
OY 617 EIFSLDEVDLP-KPSEHNHSFWNINISCSPTSGKOLLFESVTLTRPTVDLTVILAAVG 675
DB 493 KIKKLQEDDLARRKRLALYDNLWNGSLGFLPYLLPSLGRPL----- 532
OY 676 GGVLLSALGLLICVYKKKKKKTKNGPRAVGIYNGINTENPRQPKK 721
DB 533 -GILFLTLG---PCIRKTLTRIHHDKIQSGKNPRISPAPQATPNR 574
RESULT 5
YHVS_YEAST
ID YHVS_YEAST STANDARD; PRT; 1228 AA.
AC P38851;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Hypothetical 143.6 kDa protein in SPO16-REC104 intergenic region.
GN YHR155W.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=S288C / AB972;
RC MEDLINE=94378003; Pubmed=8091229;
RA Johnson M., Andrews S., Brinkman R., Cooper J., Ding H., Doyer J.,
RA Du Z., Favell A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA Kucada T., Hiller L., Jier M., Johnston L., Langston Y., Mouser L.,
RA Latteille P., Louis E.J., Macri C., Mardis E., Meneses S., Vaughan K.,
RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
RA Vaudin M.;
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT VIII"; 265:2077-2082(1994).
RL Science. 265:2077-2082(1994).
CC -1- SIMILARITY: STRONG, TO YEAST SIP3, SOME, TO S.POMBE SPAC19A8.02.
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U01397; AAB68977.1; -
CC PIR: S46754; S46754.
CC SGD: S0001198; YHR155W.
CC InterPro: IPR001849; PH.
CC Pfam: PF00169; PH: 1.
CC SMART: SM00233; PH: 1.
CC PROSITE: PS00003; PH_DOMAIN: 1.
CC Hypothetical protein.
FT DOMAIN 308 421 PH.
SQ SEQUENCE 1228 AA; 143583 MW; C8872EAE8270A4B6 CRC64;
Query Match
Best Local Similarity 2.6%; Score 114; DB 1; Length 1228;
Matches 99; Conservative 85; Mismatches 196; Indels 148; Gaps 21;
OY 179 SNGTVSRIRKQEVKALHMFHPHNVSGFSIANSSIKRLCTIES---VPEGGSAT 234
DB 301 SNGNSEEKLSG-----WLYMKTYYGHDKRYVWVARMKFLQNNFVGLSPSKT 352
OY 235 LMSANTPEGFPEDELMTQFVPAHLRAVSFLNLSNCRKEEREVEYIIPGSTNPEV 294
DB 353 YVEETDKFG-----ILMTIV-----EYLP----- 371
OY 295 FKLEDKQPGMAGNFNLSDGCDQDQSPGILRLQFOVLVQHPONESNKIVYVDLSNERA 354
DB 372 -----KEPRNFCFKRLQNPNCNCTEENTYI-----DIIQASIDELKSWINTLTSHKR 421
OY 355 MSLLI-----EPRVQOSRKRVPCGFCVLESRTCSNLTLSGSKHISFLCDDTLRLMN 410
DB 422 IALSTKEENDPKYQLARKKIEPQFFEPASSSTSTDLTLSSSKTLTF-LVEELKANNYS 480
OY 411 VEKTIQCTDHRVCOQRKSYSLQVPSDILHLPELHDFSMKLLVPR----- 454
DB 481 EDDIYSLIDKKAHLRAVRSIPATQTLHAL-----FSTFLSVNYYPCATQATWGTAWN 536
OY 455 DRLSLVLPAAQLQOHTHEKPCNTSFYLVASAIPS-----ODLYGSGFCPGSGIKO 506
DB 537 NDLSTLVNPLKGSVH---KPAIVNSNSRFSVSPDYPSLKVVDIDIOFISFISVNHDF 593
OY 507 IOYKONISVTLR---TFAPSFQOE-ASROGLTVSFTPYKEBEVFTVTP-----DKRSKY 557
DB 594 LQVPEKL-VLLRYSVWCPNNKKQFASMAFVTLNHIYVYINISGFSYLRIDLDLDSIE 652
OY 558 YLRTP-----NMDRGLPSLTSVSNISV-----PRDQVA 586
DB 653 YDKSPKIVSSRMHLHQRG-----DGLRPNMSVFTTDRRAVASKIQLFIENKAMHIIRGEKE 708


```

Db 1295 KMTKHEDYSTSSGRNNLENSDSMS-----TSGPVYIHKGDSDL----- 1334
Qy 353 RAMSLTEPRPVKQSKKEPVGCFCLESTRTCSNLTLTSGSKHKISFLCDLTRL-WMNV 411
Db 1335 -----PADQGSK-----CPECTQVAREENTO-IKENIS-----DLTCLTEIMKA 1372
Qy 412 EKTISTDHRVGOQRKSYSLQVPSDLHLVELHDFSM----- 448
Db 1373 EET-----CKKSSDKKQLPSPDKMEQNIKEFNISFQASGKNTRYSESLKSVNIN 1424
Qy 449 ----KLIVPKDRL-SLVLPAPAKLQOHT--HEKPCNTSFSYLVAISAIPQDLYFGSGFCG 501
Db 1425 RENDELTLVSDLSNKLINGINKDKMHTSCHK-----AISTKKVEEDHF-Pi 1471
Qy 502 GSRTQIQVQONISVTLRTAPSFQOASNOGLTVSF-----IPYFKE--EGVFTVTPD 552
Db 1472 VVTSQLPAAQH-----PEYEIESTKEPTLLSFHTASGKKVKIMQESLIDVKMLFDE 1522
Qy 553 TSKSVYLR-TPMWDRGLPSLTYSNMNISVPRQVACTLTPFKERSGVVCOGRAFMIIQEQ 611
Db 1523 TO---YVRKTFASGSGSKPL-----KDSKKELTLYAEK-----IEVT 1556
Qy 612 RTFAEET---FSLDEVDLPKPSFHHHSFWVNISNCSPTSGK-QLDILFVTLTPRTVDTL 667
Db 1557 ASKCEMOMQNPVSKETEMLPQONHMYROTENLKTSGTSGKQVENIENNENKMP----- 1611
Qy 668 VILLIAVGGVLLLSALGLITCCVKKKKKTKNGKPAVGIVYNGNINTEMPROPKFKGRK 727
Db 1612 -----ICICQSSYPVEDSALAYTEDSARCTVRE--SLSKGRK 1650
Qy 728 -----DNDSHVAVIETDPMY--GHLLQDSSGSFLQPVDPYRPRQGMGCPPS 775
Db 1651 WLREGQDKLGTNRTNITIECVKHEPTDFAGNASTESHVLIITREIDTNHVSQVSTL-IS 1709
Qy 776 PPTIC 780
Db 1710 DPWVC 1714

```

```

RT "Structure and chromosomal localization of the gene for the
RT oligodendrocyte-myelin glycoprotein."
RT J. Cell Biol. 111:2673-2679(1990).
RN [4]
RN SEQUENCE FROM N.A.
RN TISSUE=Brain;
RN Strausberg R.;
RN Submitted (DEC-2001) to the EMBL/Genbank/DBJ databases.
RN [5]
RN SEQUENCE OF 25-52.
RN MEDLINE=88198371; PubMed=3283151;
RN Mikol D.D., Stefansson K.;
RN "A phosphatidylinositol-linked peanut agglutinin-binding glycoprotein
RN in central nervous system myelin and on oligodendrocytes."
RN J. Cell Biol. 106:1273-1279(1988).
RN CC -1- FUNCTION: CELL ADHESION MOLECULE CONTRIBUTING TO THE INTERACTIVE
RN CC PROCESS REQUIRED FOR MYELINATION IN THE CENTRAL NERVOUS SYSTEM.
RN CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
RN CC -1- TISSUE SPECIFICITY: OLIGODENDROCYTES AND MYELIN OF THE CENTRAL
RN CC NERVOUS SYSTEM.
RN CC -1- PTM: O-GLYCOSYLATED IN ITS SER/THR-RICH REPEAT DOMAIN (POTENTIAL).
RN CC -1- SIMILARITY: CONTAINS 8 LEUCINE-RICH REPEATS (LRR).
RN CC -1- CAUTION: DO NOT CONFUSE OLIGODENDROCYTE-MYELIN GLYCOPROTEIN (OMG)
RN CC WITH MYELIN-OLIGODENDROCYTE GLYCOPROTEIN (MOG).
RN CC -----
RN CC This SWISS-PROT entry is copyright. It is produced through a collaboration
RN CC between the Swiss Institute of Bioinformatics and the EMBL outstation
RN CC the European Bioinformatics Institute. There are no restrictions on its
RN CC use by non-profit institutions as long as its content is in no way
RN CC modified and this statement is not removed. Usage by and for commercial
RN CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
RN CC or send an email to license@sib-sib.ch).
RN CC -----
RN CC EMBL: M63623; AAA59970.1; ALT_SEQ.
RN CC EMBL: X57436; CAA40684.1; -.
RN CC EMBL: X51694; CAA35991.1; ALT_SEQ.
RN CC EMBL: BC018050; AAH18050.1; -.
RN CC PIR: A30187; A30187.
RN CC PIR: A34210; A34210.
RN CC PIR: A36688; A36688.
RN CC PIR: A39613; A39613.
RN CC Genew; HGNC:8135; OMG.
RN CC MIM: 164345; -.
RN CC InterPro: IPR001611; LRR.
RN CC InterPro: IPR000372; LRR_Nterm.
RN CC InterPro: IPR003592; LRR_out.
RN CC InterPro: IPR003591; LRR_typ.
RN CC Pfam: PF00560; LRR_6.
RN CC Pfam: PF01462; LRRNT_1.
RN CC PRINTS: PR00019; LEURICHPPT.
RN CC SMART: SM00370; LRR_1.
RN CC SMART: SM00013; LRRNT_1.
RN CC SMART: SM00369; LRR_typ; 2.
RN CC KMGlycoprotein: Myelin; Cell adhesion; Repeat; Leucine-rich repeat;
RN CC signal; GPI-anchor.
RN CC SIGNAL 1 24
RN CC CHAIN 25 440
RN CC REPEAT 33 54
RN CC REPEAT 55 77
RN CC REPEAT 78 100
RN CC REPEAT 122 145
RN CC REPEAT 146 168
RN CC REPEAT 170 189
RN CC REPEAT 190 213
RN CC REPEAT 217 228
RN CC REPEAT 229 270
RN CC REPEAT 271 292
RN CC REPEAT 293 335
RN CC REPEAT 336 377
RN CC REPEAT 378 416
RN CC REPEAT 45 61
RN CC CARBOHYD 61 61
RN CC CARBOHYD 103 103
RN CC OLIGODENDROCYTE-MYELIN GLYCOPROTEIN.
RN CC LRR 1.
RN CC LRR 2.
RN CC LRR 3.
RN CC LRR 4.
RN CC LRR 5.
RN CC LRR 6.
RN CC LRR 7.
RN CC LRR 8 (INCOMPLETE).
RN CC SER/THR-RICH.
RN CC SER/THR-RICH.
RN CC SER/THR-RICH.
RN CC SER/THR-RICH.
RN CC SER/THR-RICH.
RN CC SER/THR-RICH.
RN CC SER/THR-RICH.
RN CC SER/THR-RICH.
RN CC N-LINKED (GLCNAC. . .) (POTENTIAL).
RN CC N-LINKED (GLCNAC. . .) (POTENTIAL).
RN CC N-LINKED (GLCNAC. . .) (POTENTIAL).
RN CC N-LINKED (GLCNAC. . .) (POTENTIAL).

```



```
FT CARBOHYD 152 152 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 176 176 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 189 189 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 192 192 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 364 364 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 389 389 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 425 425 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 440 AA; 49608 MW; 762FD0E8905EA050 CRC64;

Query Match
Best Local Similarity 19.1%; Score 112.5; DB 1; Length 440;
Matches 79; Conservative 62; Mismatches 144; Indels 129; Gaps 17;

QY 256 VPAHLRASVSFLNF-----NLSCERKEERVE--YYIPGTTNPEV-- 294
D 49 LPSLOENIHLNLVNHFDLHNLQYTLRLTLDISNNLESLPAHLPRSLMNMGAAN 108
QY 295 --FPLEDKOPGNMAGNFNLSLGGCDQ-----AOSPGLRLQFOVLYVHPONES 341
D 109 NNILDLK--SDYAYQWNLKYLDVSKNMLEKVLIKNTLRSLLEVNLNLSNKLMTVPPTNMP 166
QY 342 NKIYVVDLSNERAMSLTIEPRPVKSRFVGCFCVLESRTCSNLTLSGSKRKISFLC 401
D 167 SKLHIVDLSNN--SLT-----QILPGTLI-----NLT----- 191
QY 402 DDLRLMANNVKTISCSDHRYCQKRSYSLOVPSDILHLPVELHDFSKLVPKDRLSVL 461
D 192 -NLTHLILHNKKFTFIPDQSEFQ--LFQIQ-----ELTLYNRMMS--CDHKQNTIYLL 239
QY 462 VPAQKLGQHTHEKPCNTSEFYLVA-SALPISDLYFGSFCGSGSKQOYQONISVTLR-- 518
D 240 KMMETKAHVIIGTPCSQISSLKEHNMYPPPSGTSLSLFTVSGMQYDTJNISLSVTPK 299
QY 519 -----TFAPSFQOEASROGLTVEFIPY-----FKEE 544
D 300 VTKIPKQYRRKTEFGATLSKDTFTSTDKAFVYPEDTSTETINSHMAAATLTILHLD 359
QY 545 GVTF-----VTPTKSKVYLRTPNDRGLPSLTYSWNISVPROVACLTFFEKERS 595
D 360 GWTNITSLTSTKSSPTPMTLSITSGMPN-----NFSEMPQOSTTILNMEET 407

RESULT 8
PK2_YEAST STANDARD; PRT; 818 AA.
ID PK2_YEAST STANDARD; PRT; 818 AA.
AC P47116;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Serine/threonine-protein kinase PK2/STK2 (EC 2.7.1.-).
GN PK2 OR STK2 OR YJ059W OR J1725.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxId=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97079202; Pubmed=8920934;
RA Nozaki T., Nishimura K., Michael A.J., Maruyama T., Kakimura Y.,
RA Igarashi K.;
RT "A second gene encoding a putative serine/threonine protein kinase
RT which enhances spermidine uptake in Saccharomyces cerevisiae.";
RL Biochem. Biophys. Res. Commun. 228:452-458(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96457976; Pubmed=8840504;
RA Huang M.-E., Manus V., Chuat J.-C., Gallibert F.;
RT "Analysis of a 62 kb DNA sequence of chromosome X reveals 36 open
RT reading frames and a gene cluster with a counterpart on chromosome
RT XI.";
RL Yeast 12:869-875(1996).
```

```
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=97299648; pubmed=9154797;
RA Kaouss M., Audette M., Ramotar D., Verma S., de Montigny D.,
RA Gamache I., Torossian K., Poulin R.;
RT "The STK2 gene, which encodes a putative Ser/Thr protein kinase, is
RT required for high-affinity spermidine transport in Saccharomyces
RT cerevisiae.";
RL Mol. Cell. Biol. 17:2994-3004(1997).
CC -1- FUNCTION: ESSENTIAL DETERMINANT FOR HIGH-AFFINITY SPERMIDINE
CC TRANSPORT.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1sb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; D87274; BA13325.1; -
CC EMBL; 249559; CAAB9587.1; -
CC EMBL; L47993; AAB39285.1; -
CC SCD; S0003820; PK2.
CC InterPro: IPR000719; Euk_pkinase.
CC InterPro: IPR002290; Ser_thr_pkinase.
CC Pfam: PF00069; pkinase; 1.
CC ProDom: PD000001; Euk_pkinase; 1.
CC PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
CC PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
CC Transferrase; Serine/threonine-protein kinase; ATP-binding.
CC DOMAIN 255 562
CC NP_BIND 261 269 ATP (BY SIMILARITY).
CC BINDING 285 285 ATP (BY SIMILARITY).
CC ACT_SITE 388 388 BY SIMILARITY.
SQ SEQUENCE 818 AA; 91400 MW; 55B6C862894FB4BD CRC64;

Query Match
Best Local Similarity 19.2%; Score 112.5; DB 1; Length 818;
Matches 172; Conservative 127; Mismatches 317; Indels 281; Gaps 50;

QY 100 MSGPCPGEYQLOPSTSLTLNRTFIWDYKAH---KSLGLELOFSIPRLQIGPGSGCP 156
D 1 MAGNCKDEVDKSPSVSTKLKGR-LFNSSHTDSSLISAE-----QLNGRSLR 52
QY 157 DGVTN-SISG-----RIDATVVR-----IGTFCNGTV-----SRIMOE 190
D 53 KRPTSPSTISGSGGSGNSAGARQRSASLHRKKNASVG--FSNGSVSSHKSSVALQD 110
QY 191 GYKMAHLMPWFH-PRNV-SGFSIANRSSIKRLCIIEVEFGGSATYMSANYPEGPE- 246
D 111 LKHN-NNPPLNSDILGVTGSIASRDDR-AVLDRKEKEKARAKERTHTHAGLPOR 168
QY 247 -DELMTQF-----VPAHLRASVSF--LNFNLSCERKEERVEYI----- 284
D 169 SNSMSASHHFPENIVNYPYGISPNHARPDFAFDLTLTN-----KENDLSFYHMDGNSK 222
QY 285 ---IPGSTPEVFKLEDKOPGNMAGNFNLSLGGCDQDAOSPGLRLQFOVLYVHPONES 341
D 223 IRMLPFLIAPNDLPEDMKQYVHLTDNFPVTDNKPISGGSSEVR-KVKSYSRKD- 280
QY 342 NKIYVVDLSNERAMSLTIEPRPVKSRFVGCFCVLESRTCSNLTLSGSKRKI---- 397
D 281 --YVAL-----KLNMYHSPKFTYR-----CSKEFLIAHLSHNVITN 320
QY 398 -----SFLCD---DLRL-----WMNVKTICTSDHRYCQKRSYSIQ 431
D 321 TFYLKVPPTTTTTRGWFIMELGVDFOLMBRTGKKNV-----FNEKYCLEKQYAOG 375
QY 432 V-----PSDILHP---VELHDF---SKLLVPPDRLS----- 458
```

D	b		376	IKCFHNDNIARDLKREVNVLSSKEICKLDTDEGISDWHVPHDTSVVKTCQMGISPP	435
O	y	459	----	LVLVPAOKLOQHTHERPCN--TFSFY-----LVASAIPOSLLYFSGFCPGS	503
D	b	436	YTPEWVFADAKKHYPKFKOPYNPLADSVALGIMLTMINIIP-----FIDSC----	486	
O	y	504	IKOIOVKONISVTLTFAPSFOQELSROGLVSFIPTPKKEGVFLVTDPITSKYVLRPN	563	
D	b	487	-----NTDAREFEFVSVDNFNIINH-----PHRDKGCCHKPGCGEYSILARNFN	532	
O	y	564	MDRGLPSLTYSWMNISVR-----DOVACTFFEKERSGVVCQGAFMIIOERTPAE	616	
D	b	533	TD-----ATRAMRLADNPATRYMDDLFPDFEQOLETCVEPPDDOLVAPELR-KST	586	
O	y	617	EIFSLDEDVLPKPSEFHHSFWWNISNCSPSGKQLDLFSYTLTPRTVLDVILLAAVG	676	
D	b	587	STNDFENSELDAP---HDQEVIYHTSN--PLIKE-----TLTSKRGS-----MEIASP	631	
O	y	677	GVLLSALGLITCCYKKKKKTNGKPRANGIYNGINATEMPROPKR-----FOKGRD-	728	
D	b	632	SLKOKS-----KVXDSAKTKHADGDDEGNSTPKODCKENLKDEVKMGDXDK	682	
O	y	729	-----NDSHVAVIE-----DTMYGHLLDSSSGSLFPEVDTYRPFGTGMCVCPSPPT	778	
D	b	683	VIEETTTNVDSILEKPPPTSTKYEDNLSEDDS-----TMKELSMLSSTPTTP--	731	
O	y	779	ICSRAPT---AKLATEEPSPRS---PESESEPFTFSHPNNGDVSKKDTPIDLNT	828	
D	b	732	-THNGPFPLPAKAGTQLDKRMSDLJSKETASTKNFAPN---VSSSSNSLSRSLGS	784	

RESULT 9

RWL_HUMAN	STANDARD;	PRT; 1805 AA.
ID	RWL_HUMAN	
AC	O92545;	
DT	16-OCT-2001 (Rel. 40, Created)	
DT	16-OCT-2001 (Rel. 40, Last sequence update)	
DT	15-JUN-2002 (Rel. 41, Last annotation update)	
DE	Rwl protein (Fragment).	
GN	Rwl OR KIA0257.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OX	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
OX	NCLTaxid:9606;	
RN	(11)	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=Bone marrow;	
RX	MEDLINE=97191544; PubMed=9039502;	
RA	Nagase T., Seki N., Ishikawa K.-I., Ohita M., Kawarabayasi Y,	
RA	Ohara O., Tanaka A., Kotani H., Miyajima N., Nomura N.;	
RT	"Prediction of the coding sequences of unidentified human genes. VI.	
RT	The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by	
RT	analysis of cDNA clones from cell line KG-1 and brain.";	
RL	DNA Res. 3:321-329(1996).	

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>).
or send an email to license@isb-sib.ch.

CC EMBL: D87446; BAA13387.1; -.
DR BEST LOCAL SIMILARITY 19.6%; Pred. No. 5.2;
FT SEQUENCE 1805 AA; 197588 MW; 955589214413AB84 CRC64;

Query Match 2.6% Score 112.5; DB 1; Length 1805;
Best Local Similarity 19.6%; Pred. No. 5.2;
Matches 185; Conservative 117; Mismatches 363; Indels 281; Gaps 49;

38 SNITWLIKGPPTLA-KPCIYIVISKRIHTMLSIKSGERIVF-----TFSCQS--84

Dh	485	SNIFALINSPRIELAKSHHIIIGDGIETLVAEANGNTTIIISLPEEKSSLDQSSV	544
Qy	85	-PENHF-----VIEIQKNIDCSGP---CPFEVQLOPSTSLP-	119
Dh	545	TLASGYFAVEFVRKLTAKLEGIHGDAIQIOTTTDYDEILTIPIKAVIAGSLTCEPKKHVILP	604
Qy	120	-TLN--RTIIMVKAH--KSIGLEQESIPRLRIQIGPESCPDGVTHISGR	166
Dh	605	SFPGKIVHQSILNINMSFSQAKYIQIQUISLSEDFRYFKRLR--GNKEDLEPKKSKIANI	662
Qy	167	IDATVVRIGTECSNG-----TVSRIKMOEGVKMALHL---PW-FHPRNVSGFSIANRSSIK	218
Dh	663	YFDGGLCGDHCIYGLPFLPSKSEPKVQPGYAMODEMDADMDLHQSLEFGWTGIGENSGH	722
Qy	219	RLCIIESVEFEGESGATLMSANPEGPEDELTMTQEVV---PAHLRASVFLNENLSNCR	276
Dh	723	RL---SAIFEVN---TDLQNIITSIKIAE--LWSPILSLPRLIKRPL-----NTNCS	769
Qy	277	KEE-----RVETIIPGSTINPEVYKLEDKQPGMAGNFNISLOGCDQDQASPGI	325
Dh	770	EEETILENPADVVPYVQFIPILALYSNPSPV--VDK---LVSRNFNS-----KYAKIDL	817
Qy	326	LRLOFOYL--VQHQNQNSNKTIYVUDLSENRAMSLTEPRVQKSR--KVPQGFVCLSEST	382
Dh	818	RTLLEOVFRNSAHL--OSSGTFMGSLRHILILILKPGKSKVYKKEP-----VHNPT	871
Qy	383	CS--NLT-----TSGSKHISFLCDLITRLMANNVKTI--SCTDHRYCQ	424
Dh	872	VSSILIVRNNTLVMDAYWOGQGTTELRLRAGLLPBGSSLRKTIKTEALKKCTD-----	926
Qy	425	RKSYSLOVPSDILHLYELHDFSKLLVPKDRLSLVLPQAKIQOHTHEKPCNTSPSYLV	484
Dh	927	-----SLKLEPPTLKTTFKVENTGQLIHETTEIS-----	959
Qy	485	ASAIRPQDLVYGCSPGSGSFKQIOVQV--NISVY-----LRTAPRSQASASQGLVSPFI	538
Dh	960	-----GYSCEGYGFRVNNCOEFTLSANASRDIIILFTPDF--TASVIRBELKFI	1006
Qy	539	PYFKEGVFTVTPRTKSKVY-----LRTPNMDRGLPSLTSVWNISVPRDQVACLTFEK	592
Dh	1007	TTSQSEFETILNABLRPHMLATCALPRNWE-----LALYI	1044
Qy	593	ERSGVYCOTGAFNMIQEOQRTAEELFSLDEDLVLPKSPFHNSFWVNISCSPTSGKOLD	652
Dh	1045	IITSGIM--SALEFLITIGTAVLEAGIV-----EPFRRLRSF--EASNPPFDVGRPDF	1092
Qy	653	LLFVTLTRPRVDTLVL-----IAANGGVLLLSNALGLITC--CYKKKKKTKNGCP	702
Dh	1093	LRLITVJISSEG--NNTLISCDPGHSRGFCGAGGSSSRPSAGSHKQCPSPVPHSHSNRNS	1155
Qy	703	A--VGIVGNINTEMPROPKFKFGKGRKDNQSHVYAYLEIDPMYGHLL-----QDSG	752
Dh	1152	ADVENVRAKNASSSTSSSTFSQAQAASSGQANITSPVLIDSTNIVQGHNAGKSKGAKAQSQHG	1211
Qy	753	SFLQPEVDVTRPFGTGVGVCPRSPPTICSRAPTAKLATEBRPRSPSESESPYTFSHPN	812
Dh	1212	S-----QHHAHSPLE--QHQPRLPRPV-----PQOQEQPRLSLPAPLAPS	1252
Qy	813	NGD-----VSSKDTI-----PLN--TQERMER	834
Dh	1253	HPRASSARHSSSESDITSLIEAMDKDFDHHDSPALEVTELEPPSP	1298
RESULT 10			
ID	SN_MOUSE	STANDARD;	PRT; 1694 AA.
AC	Q62230; Q62216; Q62228; Q62229;		
DT	30-MAY-2000 (Rel. 39, Created)		
DT	30-MAY-2000 (Rel. 39, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Siloadhesin precursor (Stalle acid binding Ig-1-like lectin-1) (Siglec		
DE	1) (Sheep erythrocyte receptor) (SER).		
SN	OR SA.		


```

FT CONFLICT 1065 1065 S -> O (IN REF. 1; AA SEQUENCE).
SQ SEQUENCE 1694 AA; 183087 MW; 1D2BFB39B7299C9C CRC64;

Query Match
Best Local Similarity 20.6%; Score 112; DB 1; Length 1694;
Matches 109; Conservative 64; Mismatches 191; Indels 166; Gaps 25;

OY 230 EGSATLMSANYPEGPEDELMTW-----QFVPAHLR-ASVFLNFMLSNCKEKEEV 281
DB 903 EQQAVLVSCVPTGVSEGTYSYMYODGRPLQESTSTLIALISLQAGAYHCQAAPT 962
OY 282 EY-----YIPGSTNPEVFKLEKQPGNAGNFN-----LS 312
DB 963 ATASLAAPVSLHVSYPRAVITLSTPERIGHLYCVSQSPAPQQLQFHRNLVAST 1022
OY 313 LQCCDDAOS-----PGILRLOFVLYVHPQNESKRYVVDLSN----- 351
DB 1023 LQGADELASGNRLHVTVPNELRLQ-----IHPELEDDGYTCEASNTLGQASAAAPD 1078
OY 352 ERAMSTLEPRPVKQSRKFPVPGFVLESRTCSNLTLSGSKHKISFLCDDLTRW---- 408
DB 1079 AQAVRYTPVPMATVDEGQV-----NLTCLVWSTHO-----DSLSTWYKG 1119
OY 409 -----MNVEKTISTDHRVC-----QRKSYSLQVPSDILHLPELHDFSMKLL 451
DB 1120 GQQLGARSTILPSVAVLATSYR-CGVGLPGHAPLSRPVTLDVLAHARNLR-LTYLLE 1177
OY 452 VKRDLSTLVV-----PAQKLOQHTHERPCMTSFSYVAS---AIP----- 489
DB 1178 TQGRQIALVLCVDSRRPQQLTSLHGDQ-----LVASSTEASVPMTLRLQLDPPR 1228
OY 490 -GODLY-FESFCPGS-----IKQIQVKNISVLTLPFAPSFOEASRQGLTSPFLP 539
DB 1229 SNEGSLCSAHSPLGANTSLLELLEGVAKNMPGSGSVEGEVYVTCEDPALLSALYA 1288
OY 540 YFK-----EFG-----VFTVPTDKSKVYLRTPMNDGRPSLTSYSMNI-SVPRDVA 588
DB 1289 WFNHNGHWLOEGPASSLOFLVTRAHGAATFCQVHDTQGRSSRPAQLQTLAPRD--AVL 1346
OY 589 TFFKERSGVVCGTGAEMTIQ---EORTRAELIFSLDEVDLPKPSFHNS 635
DB 1347 SSFRD-----SKTRLVAVIQCTVDEPPEAEVLSHGKVLASHERHSS 1390

RESULT 11
TRKL_LYMYST STANDARD: PRT; 794 AA.
AC 076997;
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative neurotrophin receptor LTRK 1 precursor (EC 2.7.1.112).
OS Lymnaea stagnalis (Great pond snail).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Basommatophora;
OC Lymnaeidae; Lymnaea.
OX NCBI_TaxID=6523;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=9823249; PubMed=9564036;
RA van Keesteren R.E., Fainzilber M., Hauser G., van Minnen J.,
RA Vreugdenhil E., Smit A.B., Ibanez C.F., Gereerts W.P.M.,
RA Bulloch A.G.M.;
RT "Early evolutionary origin of the neurotrophin receptor family.";
RL EMBL J. 17:2534-2542(1998).
CC -1- FUNCTION: MAY BIND AN ENDOGENOUS INVERTEBRATE NEUROTROPHIN. BINDS
CC HUMAN NT-3, BUT NOT NGF OR BDNF.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine - ADP + protein
CC tyrosine phosphate.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: EXPRESSION IS CONFINED TO THE CENTRAL NERVOUS
CC SYSTEM AND ITS ASSOCIATED ENDOCRINE TISSUES.
CC -1- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES. INSULIN
CC RECEPTOR SUBFAMILY.

```

```

CC -1- SIMILARITY: CONTAINS 2 LEUCINE-RICH REPEATS (LRR).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
CC EMBL; 061728; AAC26840.1; -.
CC HSSP; P08631; 1AD5.
DB InterPro: IPR000719; Euk_pkinase.
DB InterPro: IPR001611; LRR.
DB InterPro: IPR000372; LRR_Nterm.
DB InterPro: IPR003591; LRR_Typ.
DB InterPro: IPR002011; RtkkinaseII.
DB InterPro: IPR001245; Tyr_pkinase.
DB Pfam; PF00069; pkinase; 1.
DB Pfam; PF00560; LRR; 1.
DB PRINTS; PR00109; TYRKINASE.
DB ProDom; PD000001; LRR_pkinase; 1.
DB SMART; SM00013; LRRNT; 1.
DB SMART; SM00369; LRR_Typ; 1.
DB SMART; SM00219; TYRK; 1.
DB PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DB PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DB PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DB PROSITE; PS00239; RECEPTOR_TYR_KIN_II; 1.
DB Transfaser; Tyrosine-protein kinase; Transmembrane; ATP-binding;
DB Phosphorylation; Receptor; Glycoprotein; Neurogenesis;
KW Leucine-rich repeat; Repeat; Signal.
FT SIGNAL 1 33
FT CHAIN 34 794
FT DOMAIN 34 419 POTATIVE NEUROTROPHIN RECEPTOR LTRK 1.
FT TRANSMEM 420 440 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 441 794 POTENTIAL.
FT REPEAT 161 204 CYTOPLASMIC (POTENTIAL).
FT REPEAT 205 228 LRR 1.
FT DOMAIN 504 775 LRR 2.
FT NP_BIND 510 518 ATP (BY SIMILARITY).
FT BINDING 538 538 ATP (BY SIMILARITY).
FT ACT_SITE 647 647 BY SIMILARITY.
FT CARBOHYD 64 64 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 102 102 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 128 128 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 288 288 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 374 374 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT MOD_RES 673 673 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 677 677 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 678 678 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 789 789 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ SEQUENCE 794 AA; 89054 MW; FFE3EF576E1A440 CRC64;

Query Match
Best Local Similarity 20.7%; Score 111; DB 1; Length 794;
Matches 119; Conservative 62; Mismatches 182; Indels 212; Gaps 32;

OY 203 PRNVSGFSIANRSSIKRLIIESVFEGESATLMSANYPEGPEDELMTWQFVPAHL-- 260
DB 4 PRFRRLTRANVLTV--ISILSILSGAGCSPLSQ-----LPSDN-----PAHGV 47
OY 261 -----RASVFLNFMLSNCKEKEVEYIIPGSTNPEVFKLEDKOPN 304
DB 48 ODGVTERVDRSKNHNHTTASSGAHRTVSGEPGDEV---TTRSTVAP-----DQVPG 98
OY 305 -----MAG-----NNLSLQGGDDAQ-----SGILRLQFVLYVHPQNE 340
DB 99 ASRNTTMAGTKCSLOYDLSTFPACQACQANSEGVAVSCVPTDLR-EPVAREVARA 157
OY 341 SNRIYVVDLSNERAMSLTIEPRPVKQSRKFPVPGFVLESRT---CSSN-----LTLT 390
DB 158 -----YIKLELRQOSKLT-----SLKTEIKF-----FICLHILTIENGCLNIGSIARKTIL 204

```

```

QY 391 S-----GSHKISFLCD-----DLRLMNVK-TISCIDHRXCOQRYSLOVPSDILH 438
DB 205 SLETLNRHNLHLEFPDELLRLTLRLMELGNALFCSCTN-----LMLRSVDVAADRS 259
QY 439 LPELHNFMSKLVLPKDRSLVLPKAOLOQHHEKPCN-TSPSYLVASAPSDLYFGS 497
DB 260 MCGSTRG-----VSKMKMFOFCE-PCGIPDIRNMMLVEPKNGMFLRL 303
QY 498 FCPGSGIK-OIOVKONISVTLRTFAPSFQOEASRGLTVSFIPFKE-----GVFTV 549
DB 304 FVLSGCKPKRIDLRHNLHNRSSGSSQF-----LIDKSEFNGQVYTGITTI 351
QY 550 TP--DFRSKYVLRTPMWDRLPSLTVSWMISVPRQVACLTFEKRSGVVCOTGRAFM 607
DB 352 LPHNETSQTYY-----LTAVN----- 368
QY 608 IQORRAEIEFSLDEVDLPKPSFHHSFVNISNCSPTSGKOLDLFVTLPR--TV 664
DB 369 ---SKGANQTFHLVDQTPASSIH-----IPLSNIPR-----ISSATTPRASPT 412
QY 665 DL---TVILAAVGGVLLSALGLITCYKRRKK 696
DB 413 DFGPQVITLPVGVVILLISAVFIYLCQRAKHR 447

RESULT 12
CRAR_MOUSE STANDARD: PRT: 704 AA.
ID CRAR_MOUSE
AC P98054;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Complement-activating component of Ra-reactive factor precursor
DE (EC 3.4.21.-) (Ra-reactive factor serine protease p100) (hARF)
DE (Mann-binding lectin serine protease 1).
GN MASPI OR CRARF.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10990;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-BALB/C; TISSUE-Liver;
RA Takayama Y., Takada F., Takahashi A., Kawakami M.;
RA MEDLINE=94179811; PubMed=8133044;
RT "A 100-kDa protein in the C4-activating component of Ra-reactive
RT factor is a new serine protease having module organization similar to
RT C1r and C1s."
RL J. Immunol. 152:2308-2316(1994).
RN [2]
RP SEQUENCE OF 465-704 FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-BALB/C; TISSUE-Liver;
RA MEDLINE=93176166; PubMed=849319;
RA Takahashi A., Takayama Y., Hatusue H., Kawakami M.;
RT "Presence of a serine protease in the complement-activating component
RT of the complement-dependent bactericidal factor, hARF, in mouse
RT serum."
RL Biochem. Biophys. Res. Commun. 190:681-687(1993).
CC -1- FUNCTION: COMPONENT OF THE BACTERICIDAL RA-REACTIVE FACTOR RARF
CC WHICH SPECIFICALLY BINDS TO RA AND R2 POLYSACCHARIDES EXPRESSED BY
CC CERTAIN ENTEROBACTERIA. IT TRIGGERS THE ACTIVATION OF COMPLEMENT
CC CASCADE BY CLEAVING THE C4 AND C2 COMPONENTS. IT ACTIVATES THE
CC C4 COMPONENT BY CLEAVING THE ALPHA-CHAIN OF C4.
CC -1- SUBUNIT: RARF CONSISTS OF A COMPLEMENT-ACTIVATING COMPONENT
CC (CRARF) AND A POLYSACCHARIDE-BINDING (MANNOSE-BINDING) COMPONENT.
CC CRARF IS AN HETERODIMER OF A HEAVY (P70) AND A LIGHT CHAIN (29)
CC LINKED BY A DISULFIDE BOND.
CC -1- TISSUE SPECIFICITY: LIVER.
CC -1- DOMAIN: CRARF HAS A MOLE ORGANIZATION SIMILAR TO C1R AND C1S.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -1- SIMILARITY: CONTAINS 2 SUSHI (SCR) DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.

```

```

CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D16492; BAA03944.1; -.
DR HSSP: P00736; 1APO.
DR MEROPS: S01.198; -.
DR MGD: MGI:88492; Maspl.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000859; CUB domain.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR001254; Ser_protease_Try.
DR InterPro: IPR000436; Sush1_SCR_CCP.
DR Pfam: PF000084; sush1; 2.
DR Pfam: PF000089; trypsin; 1.
DR Pfam: PF00431; CUB; 2.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00032; CCP; 2.
DR SMART: SM00042; CUB; 2.
DR SMART: SM00179; EGF_CA; 1.
DR SMART: SM00020; Tryp_Spc; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; 1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01187; EGF_CA; 1.
DR PROSITE: PS02040; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR Hydroxylase; Complement pathway; Serine protease; Protease;
KW Glycoprotein; Sush1; Repeat; Signal; EGF-like domain; Hydroxylation.
FT FT CHAIN 1 24
FT FT 25 704
FT FT
FT FT CHAIN 25 453
FT FT 454 704
FT FT
FT FT DOMAIN 25 143
FT FT 144 187
FT FT
FT FT DOMAIN 190 302
FT FT 305 368
FT FT
FT FT DOMAIN 371 438
FT FT 454 704
FT FT
FT FT ACT_SITE 495 495
FT FT 557 557
FT FT
FT FT ACT_SITE 651 651
FT FT
FT FT MOD_RES 164 164
FT FT
FT FT DISULFID 148 162
FT FT 158 171
FT FT
FT FT DISULFID 173 186
FT FT 190 217
FT FT
FT FT DISULFID 247 265
FT FT 306 354
FT FT
FT FT DISULFID 334 367
FT FT 372 419
FT FT
FT FT DISULFID 402 437
FT FT 441 577
FT FT
FT FT DISULFID 619 636
FT FT 647 677
FT FT
FT FT CARBOHYD 34 34
FT FT 183 183
FT FT
FT FT CARBOHYD 390 390
FT FT
FT FT CARBOHYD 412 412
FT FT
SO SEQUENCE 704 AA; 79895 MW; 71F4F3012D2C67F CRC64;

```

Query Match

2.5%; Score 110.5; DB 1; Length 704;


```
Db 211 LPDIEDKQVOMFLTRYL---TNDVRMNFNFNSDM---FNPLS-----SWS 252
Oy 328 LOFOVLVQHPOHNSKIYVDSN-----ERAMSLTI 359
Db 253 LONELSNTRQOOSMSSTPIKNLGRDAQNFLLGLLNKNGWSLEDSGSVEIDI 312
Oy 360 EBRPVKQSHKEVPGCFVCLSRCS-----SNLTGSGSHKISF--LCDDLTRLMNV 411
Db 313 SGTITPGHIVYFGCMVLEGIYYSVGNKFHYTSKMLTPGGERREITLETIGNLDLGIH- 371
Oy 412 EKTICTDHRVYCGKRSYSLSQVPSDLH--LPVELHDFSMKLVPR---DRLSLVLPAPKL 467
Db 372 ---GISNNNFARLDKDLKIR---LHLEKETLHKKFVILCANLEFLDLK--IMPLSKI 423
Oy 468 QOHTHEKPCNTFSYLVASAIPOSOLYRSGSPGSGIKIOVKONISVTLRFPAPFOOE 527
Db 424 LQKLNDP-----PTLLIMQSF-----TSVPYFASM 450
Oy 528 ASRQGLTVAFIPYFKE-----GVFTVTPDTKSKVYLTPN--WDRGLPLTSVSN 577
Db 451 SSR---NISSQFKNPDLATLSRFNLTENTMTITPGNDLMG---SMVSLGAS 503
Oy 578 ISVPDQVACLTFEERSGVVCGTGRAFMIIQEOBTR---AEELFSLDEYLPKPSFH 633
Db 504 GLPDQDIP--SAFTKKIKVCKN---VWSSNPTRIAYLSOEIYIFRDLSGRKRRH 557
Oy 634 HSPWNISCSPTSGKOLDLFSVTLTPRTVDTVYLLAAGGVLLSALGLIICVAK 693
Db 558 LEFPNESDVTYENDM-----LVKTIIDOGH----- 575
Oy 694 KKKTKNGPAGVIGYNGININTEMPROPKFKGRKNDSHVAVIEDTMYGHLDDSSGS 753
Db 576 MKKDIDYPI-----DELKPEPDQLEPOKVOETRK----- 614
Oy 754 FLQPEVDYTP---FQGTWGVCP--PSPTTIC 780
Db 615 -LSPFLDSLRIPIWDLHTLTLCPPTSTMVLC 645

RESULT 15
MAP2_MOUSE STANDARD: PRT: 1828 AA.
ID MAP2_MOUSE
AC P20357:
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Microtubule-associated protein 2 (MAP 2).
GN MAP2 OR MAP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89083571; PubMed=3205744;
RT Wang D., Lewis S.A., Cowan N.J.;
RT "Complete sequence of a cDNA encoding mouse MAP2.";
RL Nucleic Acids Res. 16:11369-11370(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89043973; PubMed=3142041;
RT Lewis S.A., Wang D., Cowan N.J.;
RT "Microtubule-associated protein MAP2 shares a microtubule binding
RT motif with tau protein.";
RL Science 242:936-939(1988).
CC -1- FUNCTION: THE EXACT FUNCTION OF MAP2 IS UNKNOWN BUT MAPS MAY
CC STABILIZE THE MICROTUBULES AGAINST DEPOLYMERIZATION. THEY ALSO
CC SEEM TO HAVE A STRENGTHENING EFFECT ON MICROTUBULES.
CC -1- SIMILARITY: CONTAINS 3 TAU/MAP REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

```
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M21041; AAA39490.1; -.
DR PIR: S06467; S06467.
DR PIR: A40115; A40115.
DR MGD: MGI:97175; Map2.
DR InterPro: IPR001084; Tubulin_Tau.
DR Pfam: PF00418; Tubulin-binding.
DR PROSITE: PS00229; TAU_MAP_2.
KW Microtubules; Repeat; Calmodulin-binding.
FT DOMAIN 1452 1472 CALMODULIN-BINDING (POTENTIAL).
FT REPEAT 1662 1692 TAU/MAP MOTIF.
FT REPEAT 1693 1723 TAU/MAP MOTIF.
FT REPEAT 1724 1755 TAU/MAP MOTIF.
SQ SEQUENCE 1828 AA; 198980 MW; 200BC59E3360538CA CRC64;

Query Match 2.5%; Score 108.5; DB 1; Length 1828;
Best Local Similarity 18.1%; Pred. No. 11;
Matches 139; Conservative 108; Mismatches 285; Indels 237; Gaps 32;

Oy 228 EEGSATLMSANVPPEGFPEDELTMTQFVYPAHLRASVFLN-----FNLSNCE--RKEER 280
Db 546 EGAGSATIIEVEMP--FYEDKSGMSKYFETSALEKEDMTSTELGSDYELSDSRGSAQS 603
Oy 281 VEYIIPGSTNPVEFKLEKOPGNMAGNFNLSQGDQDAQSPGILRLQFOVLVQHPONE 340
Db 604 LDTISPKNHDEKELQAKASOPSPPAQEAGYSTLA---QSYPFG---HPSELPEPSSP 656
Oy 341 SKTIYV-----DLSNERAMSLTIEPRPVKQSKRVYPCGFVCLERTCSSNLTLS 391
Db 657 QERMFTIDPKVYGEKRDLSKKNKDLTL-----SRSLGLGRSALIEGRMSINLPMSC 709
Oy 392 -----GSKHKISFLCDDLTRLMNVKTICTDHRVYCGKRSYSLSQVPSDLH--- 438
Db 710 LDISALGFNFGCHDSLPLASDI-----LNTSGSMDGDDYLPPTT 751
Oy 439 -----LPVELHDFSMKLVPRDLSLVLPAPQRLQOHTHEKPCNTFSYLVASAPDS 490
Db 752 PAVEKMPCFPIESKEEDEKAEOAK-----VYGGQTIQVETSESPFAKEYKKNQTVMA 805
Oy 491 QDLRSGSPGSGIKIOVKONISVTLRFPAPFOEASRQGLTVSFIFYKEEGFYTY 550
Db 806 PDLPEMLDLAGTRSRASVSADAEVARRKSVPS---EAMLAESSYS-LPVADESPYVK 861
Oy 551 PDKSKVYLRTPWMDRG-----LPSLSVSNIS-----VPRDQVAC 587
Db 862 PDSQLE-----DMGYCVFNKTYVPLPSPVODSENLSGSGSFYEGTDKVRDLARD 913
Oy 588 LTFEKK-----ERSGV-----VCQTRAFM 606
Db 914 LSLIEVKLAAGRVKDEFAEKAETPTTSADKSGLSREFDHRKANDKLDYLEKSEHI 973
Oy 607 IIOEQRTAE-----ETISL-----DEVLPRPSFIHHFWMNISCSPT 646
Db 974 DSRHAKESSEEMGKVELGLGITYDOASTKELITTKDTSPEKTENGSLSVPEVAVEPT 1033
Oy 647 SGKQDLDFSVT-LPRTVDLTVYLLAAGGGVLLSALGLIICVKKKKTKNKPAAG 705
Db 1034 TKADQGLDPAATKAEPQSLDIKYSDFGQMASG-----AVDAGKALE 1075
Oy 706 I-----YNGNINTEMPROPKF-----QKGRKNDSHV-YAVIEDTMYGHLQ--- 748
Db 1076 LKFEVAQELTISSEAPQADSFVWGVSCHIKKEGKVNETEVKRYKPLDLVQEAVDKBE 1135
Oy 749 -DSSSFLOPEVDYTRPQGTGVCPPSPPTICSAAPPAKLATEPPRRSPRESSEPY 806
Db 1136 SYESSGEHSLTMSLKPDGKKETSPET--SLIDEVAKLKLVEIIPC--PPVSEAD-- 1189
Oy 807 TFSHPNKGVD-----SKNDIDPL--NFOEPM-----EPAE 836
```

Db 1190 -LSTDEKGEVOMEFIQLPKKESTETPDIPAIPSDVTQPOPEAIVSEPAE 1237

Search completed: November 11, 2002, 11:31:35
Job time : 35 secs

...